Multilevel Synthesis

Daniel Courgeau

MULTILEVEL SYNTHESIS

From the group to the individual

Daniel Courgeau
Ined (Institut national d'études démographiques), Paris, France

To Hella, my wife

Contents

Chapter III. – **Analyzing individual data**

1. Establishing event-history analysis
   - Introducing individual behaviors
   - Introducing time regressions
2. Paradigm of event-history approach
3. Problems encountered with this approach
   - Unobserved heterogeneity
   - Risk of atomistic fallacy

Chapter IV. – **Toward a contextual and multilevel analysis**

1. Establishing the analysis
   - From contextual analysis…
   - …to a multilevel analysis
2. Toward a synthetic paradigm

PART II
Multilevel analysis

Chapter V. – **Defining levels**

1. Different types of levels
   - Social or economic groups with obvious effects
   - Geographic and administrative groups with less direct effects
2. Different types of nestings
   - Period observation
   - Event-history observation

Conclusion

Chapter VI. – **Linear analysis of continuous characteristics**

1. A two-level linear-regression model
2. Estimating model parameters
3. Risks of erroneous inference

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

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

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.

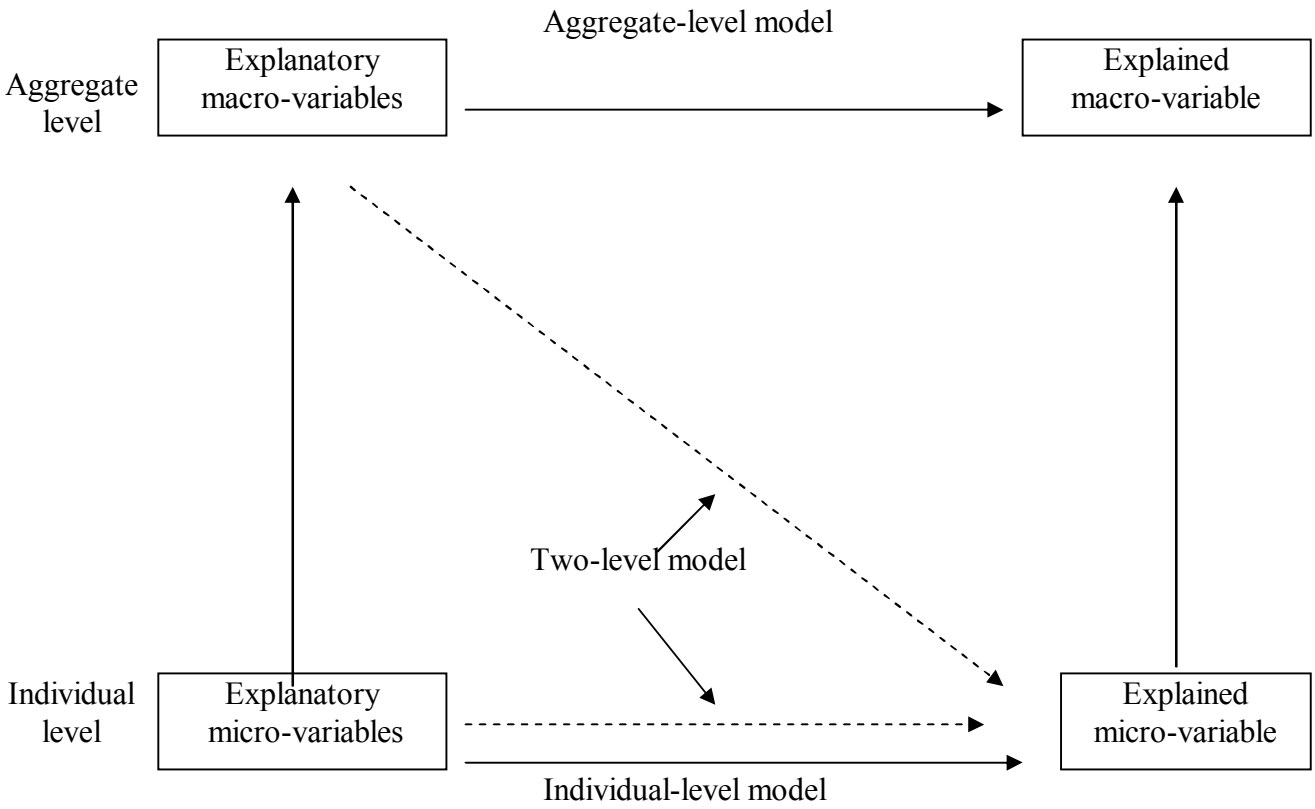
The paradigms—or, rather, research programs—that sustain such an approach in the various social sciences will now regard individual event histories (biographies) as the life experience on which they will work. However, their research objects will consist of the processes that impart meaning to the event 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.

Figure 1. - Links between different models



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 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, 2001). 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, 2002). 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, 1961; Henry, 1972; Schryock and Siegel, 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, 2003; etc.), which supply full details on the mathematical and statistical approaches used in social science.

Summary

This textbook presents a historical panorama of the evolution of demographic thought from its eighteenth-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.

PartI

From macro/micro opposition to multilevel analysis

CHAPTER I

PERIOD ANALYSIS OF SOCIAL GROUPS

In the early days of "Political Arithmetic," authors did not always draw a clear distinction between the period approach (also known as cross-sectional approach), which examines events occurring at a specific point in time, and the cohort approach (also known as generational approach), which examines the events occurring over a person's entire lifetime. By the eighteenth and nineteenth centuries, however, the period approach prevailed.

John Graunt (1662) analyzed the annual "Bills of Mortality" and "Bills of Christenings" to compile a cross-sectional picture of the population of London and the English counties. Likewise, Edmond Halley (1693), working on annual data for the city of Breslau from 1697 to 1691, compiled an acceptable life table for a population that he could regard as stationary. On the other hand, the links between annuity values and population mortality led some authors to discard the period approach in favor of a cohort approach to mortality. Following in the steps of predecessors such as Jean de Witt in 1671 (Dupâquier, 1985), Antoine Deparcieux (1746) built life tables of recipients of the 1689 and 1696 tontine annuities. He observed them until early 1742, classifying recipients by their age at the time of constitution of the annuity. In Deparcieux's day, this use of a cohort approach did not contradict the period approach, for the underlying assumption was that of a stationary population.

By the late eighteenth century, the introduction of censuses in many countries consolidated the cross-sectional approach. As a census enumerated a country's population at a specific moment, it entailed the examination of vital statistics pertaining to that same point in time, for the purpose of calculating period demographic indices. This type of analysis was practiced until the end of World War II.

This chapter begins by looking at how period analysis was established and how it functions. Next, we identify its more basic principles, which we may describe as the paradigm or research program, and which we can then discuss more specifically.

I. From a descriptive approach to a statistical period analysis

As noted above, demography initially sought to develop a cross-sectional description of demographic phenomena, by compiling population pyramids, crude rates, age-specific rates, and synthetic (i.e., overall) indices to summarize these rates. As space precludes a detailed presentation of this approach here, we shall only outline its principles.

The tools to describe period statistics were partly established by "political arithmeticians" and continuously refined during the nineteenth century and the first half of the twentieth. Some notions, such as the birth rate or the christenings/marriages ratio, soon appeared inadequate to measure population fertility. New indices were successfully introduced, raising new issues in censuses and civil registration. For example, Joseph von Körösi (1894) proposed the construction of period fertility tables by spouse age, allowing an analysis of fertility that was more complex but closer to reality. Synthetic indices—such as total fertility rates and cumulated first-marriage frequencies—were developed; later, we examine the problems that they can generate. Other improvements were suggested, such as introducing female age at marriage for legitimate fertility, children's sex and birth order, and parental religion and occupation. Adolphe Landry's *Traité de démographie* (1945) offers an excellent summary.

As soon as demographers try to go beyond a mere description of phenomena and seek to determine whether a given factor influences the phenomenon studied, they will need to define the factor, measure it more precisely, and see how they can identify—with the aid of a model—the links they want to demonstrate. Before we tackle modeling issues, let us discuss measurement issues.

As shown above, the approach adopted here requires an aggregation of measurements in order to estimate the probabilities of various events. Consequently, to estimate the likelihood of a given event, we must examine either an entire population, or sufficiently large sub-populations. But if we then want to link that event to another event or characteristic, we cannot do so by working on the entire population: two marginal probabilities will give us no information on the possible link between them. To demonstrate such a connection, we must look at sub-populations. To take a simple example, which we shall use throughout this book, if we only know the probability of mobility for the total population and its percentage of farmers, we shall be unable to demonstrate a link between these two quantities. To do so, one solution will consist in decomposing the population into a large number of sub-populations and estimating the two quantities in each. We can then measure the probabilities of migrating for farmers and the rest of the population, under certain hypotheses that—as shown later—can vary.

Clearly, this reasoning did not emerge when the approach was first used, although it is implicit in the early attempts to connect different demographic phenomena. To some extent, the studies by Johann Peter Süßmilch (1741, 1765, 1979, 1998) already moved in this direction. Comparing urban and rural mortality in several countries, he concluded:

> "The difference between towns and villages must be sought in people's eating habits, manners, and ways of life." (1979, p. 324)

This showed the relationships that can exist between various phenomena. Similarly, a century later, Adolphe Quetelet (1869) introduced his theory of the average man by noting that:

> "Man is under the influence of causes most of which are regular and periodic. We can, through close study, determine these causes and their mode of action, as well as the laws [distributions] to which they give birth; but, to do this successfully, we must study the masses, so as to rid our observations of all that is merely fortuitous or individual. Probability calculus shows that, all things being equal, we shall come that much closer to the truth or the laws that we want to grasp as the observations encompass a larger number of individuals." (p. 33)

Of the different applications of such an approach, we shall take the example that Quetelet borrows from Michael Thomas Sadler (1830). When he seeks to link fertility and nuptiality in a given country, Quetelet breaks down the country into *départements* (France) or provinces (Netherlands). For each unit, he computes the proportion of marriages and the number of legitimate births per marriage. Comparing these two quantities, he observes:

> "The places that produce the most marriages annually are those where marital fertility is lowest, through a sort of compensation that prevents a country from experiencing overly rapid increases in population." (p. 80)

Quetelet therefore effectively observes a negative correlation between these aggregated quantities, without in any way seeking to show that one causes the other. Such a correlation implies no causality criterion and may be due to a third factor, even absent any causal relationship between the first two.

In our view, it is Émile Durkheim (1895, 1897) who gave the clearest exposition of the goals of this social science and the means to attain them, while criticizing the approach of his predecessors, particularly Quetelet's "average man." The latter theory enables us to explain human behaviors, if we assume that they depend exclusively on the country in which they occur and the correlations existing between different social facts in the society. But their own origin is not elucidated, and we need to search further for their deep causes.

First, Durkheim observes that social facts are independent of their function in society and that they can serve different purposes while remaining unchanged. What is important to identify is the function; showing a fact's usefulness becomes secondary. But the function cannot be examined without taking into consideration the different constituent parts that the society under study may contain, i.e., the religious, domestic, political, and other groups associated with it. Social phenomena must vary with the forms of this association and with the way in which the constituent parts of society are grouped together. We can therefore identify the function of a social fact by linking it to other social facts, but it is the social system that underpins this explanation (Franck, 1994).

In such circumstances, how do we prove that one phenomenon is the cause of another? Ethical reasons preclude the use of the experimental approach, narrowly defined, in most social sciences. That would require dividing subjects at random into two groups, one subjected to the treatment we want to test (target group), the other being totally dispensed from it (control group). Applied to a sample of adequate size, this procedure enables us to check the parasite risk factors including unknown ones—apart from pure random events. Such an approach is

possible only when the scientific protocol does not conflict with the patients' interests. This drastically reduces the scope for such tests in demography, although some efficiency tests for new contraceptive methods have been attempted in the discipline (Wunsch, 1994).

We therefore need to use a comparative method. One of the most relevant is the method of concomitant variations, as proposed by Durkheim (1895):

> "For [the method] to be demonstrative, we do not need to strictly exclude all the variations that differ from the ones we are comparing. The mere parallel pattern of the values assumed by both phenomena, provided that the pattern has been identified in a sufficient number of sufficiently diverse cases, is the proof that a relationship exists between them. The method owes this privilege to the fact that it touches on the causal relationship." (p. 129)

The method is the same as the one proposed later by Landry (1945, p. 26), who notes that, if we want to "understand a variation across time, a difference across space," we will need to identify "a relationship, a relationship of equality, concomitance, covariation, or of any other sort" between the phenomena studied. Ultimately, this is equivalent to a regression analysis—to use today's terminology—between for example the percentages of suicides and the percentages of Protestants living in different regions.

Observing a single characteristic

Let us show how we can formulate such a model more precisely, in the case where we observe the number of individuals having experienced a given phenomenon as a function of their membership or non-membership in a population category in different areas of a country.

Let us assume that the members of the population category studied are all at equal risk of experiencing the phenomenon, whatever their territorial location: they are therefore homogeneous with respect to the phenomenon under study, and the phenomenon is unrelated to the individuals' other life events. This is an initial hypothesis, identical to Durkheim's (1895): he posited that the members of a group had the same propensity to experience a given event—such as suicide—irrespective of their province of residence. Likewise, this propensity is independent of the other social phenomena for, as quoted above, "we do not need to strictly exclude all the variations that differ from the ones we are comparing."

In this case, we shall demonstrate that, if the initial hypotheses are verified, we shall find a linear relationship between these two characteristics. Let p_1 and p_0 be the probabilities of experiencing the event, for individuals possessing the characteristic studied or not; the intensity of the probabilities is identical in all the regions of the country examined. Let $X_1(r)$ and $X_0(r)$ be the sizes of the two populations in a given region r. The mathematical expectation of the number of events observed, $E[Y(r)]$, will be equal to:

$$E[Y(r)] = X_1(r)p_1 + X_0(r)p_0 = X_1(r)(p_1 - p_0) + X(r)p_0 \qquad \text{(I.1)}$$

where $X(r)$ is the total population of region r. We effectively find that the proportion of individuals experiencing the event in each region, $y(r)$, will be a linear function of the proportion of individuals in the region who possess the characteristic studied, written more simply as $x_1(r)$, to which we need to add a random term ε_r :

$$y(r) = x_1(r)(p_1 - p_0) + p_0 + \varepsilon_r \qquad \text{(I.2)}$$

If the model's hypotheses are correct, we can estimate the parameters of this linear relationship using classic linear-regression methods.

However, it is important to realize that, in this case, the estimated parameters are in fact constrained by the fact that p_1 and p_0 are probabilities, which must fall within the interval [0,1]. Now, in a regression of this type, there is no reason why the estimated coefficients should comply with the constraint: as a result, the interpretation posited *a priori* could be totally invalidated. When discussing the methodological issues raised by this approach, we shall see in greater detail the difficulties involved in the estimation.

Example no. 1: Suicide in Prussia

Let us take a more detailed look at the example given by Durkheim (1897), who studied suicide rates per million inhabitants in the thirteen provinces of Prussia between 1883 and 1890, as a function of their percentages of Protestants (table I.1). We can treat these data in different ways. Let us first consider the mean suicide rates per province as a function of their percentage of Protestants (last line of table I.1). We obtain the points of figure 1, which allow an estimation, using a simple linear regression, of the parameters of the corresponding regression line. The regression is very satisfactory, with a 0.95 share of variance explained by the model. The parameters and their standard deviation, estimated as above, are given in the first columns of table I.2.

TABLE I.I. - SUICIDES, PER MILLION, BY PERCENTAGE OF PROTESTANTS IN PRUSSIAN PROVINCES (1883-1890)

Provinces with more than 90% of Protestants	Suicide rate, per million	Provinces with between 89% and 68% of Protestants	Suicide rate, per million	Provinces with between 40% and 50% of Protestants	Suicide rate, per million	Provinces with between 28% and 32% of Protestants	Suicide rate, per million
Saxony	309.4	Hanover	212.3	Western Prussia	123.9	Posen	96.4
Schleswig	312.9	Hesse	200.3	Silesia	260.2	Rhineland	100.3
Pomerania	171.5	Brandenburg and Berlin	296.3	Westphalia	107.5	Hohenzollern	90.1
		Eastern Prussia	171.3				
Mean	*264.6*	*Mean*	*220.0*	*Mean*	*163.6*	*Mean*	*95.6*

Source: Durkheim, 1895, p. 151.

We can thus conclude that Protestants have a 277 per million probability of committing suicide, compared with 36 per million for the other religions (essentially Catholics), i.e., 7.7 times as high. This is tantamount to saying that "suicide varies in inverse proportion to the degree of integration of religious society" (Durkheim, 1897, p. 222). Durkheim likewise establishes a relationship between suicide and domestic and political society, by examining—again separately—the effect of various family characteristics (bachelorhood/spinsterhood, marriage, widowhood, etc.) or political characteristics (effects of crises, wars, revolutions, etc.). In seeking the common cause of these different outcomes, Durkheim was moved to write:

> "The cause can only lie in an identical property that all these social groups possess, albeit, perhaps, to different degrees. Now the only property that meets this condition is that all these social groups are highly integrated. We therefore arrive at this general conclusion: suicide varies in inverse proportion to the degree of integration of the social groups to which the individual belongs." (p. 223)

It should be recalled here that this conclusion is valid if the hypotheses on which the model is based are verified, i.e., if the group behaviors are homogeneous and independent of one another.

TABLE I.2.- PARAMETERS ESTIMATED FOR THE REGRESSION MODEL GIVING SUICIDE RATES, PER MILLION, AS A FUNCTION OF THE PROPORTION OF PROTESTANTS, IN PRUSSIA (1883-1890)

Probabilities per million	Regression on means		Regression on provinces	
	Estimated value	Standard deviation	Estimated value	Standard deviation
Suicides in other religions	36.4	21.5	37.0	43.8
Suicide among Protestants	277.0	14.6	276.3	28.5
Adjusted R^2	0.95		0.52	

Source: Durkheim, 1895.

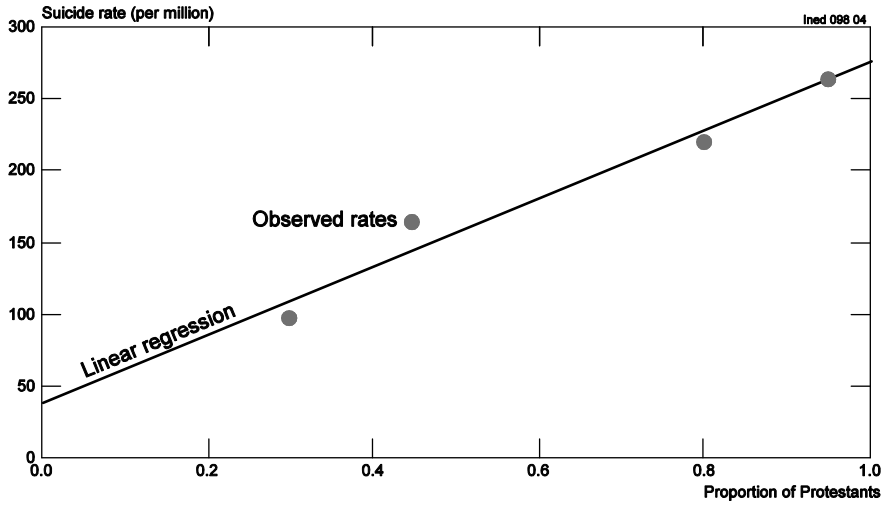


Figure I.1. - Suicide rates per million according to proportion of Protestants (Prussia, 1883-1890)

One could also estimate the same regression on the data of each of the thirteen provinces. Table I.1 tells us less about the percentages of Protestants in each province, because of their aggregation. If we assume, as before, that the provinces of each class have their mean percentage of Protestants, we can proceed with the estimation, which yields the parameters given in the last columns of table I.2. The parameters are very close to the previous ones, but the standard deviation of the estimates is twice as large and the share of variance explained by the model is far lower at 0.52. These figures show the model's strong dependence on the number of regions examined. Later, we shall see the type of error generated by such a model.

Example no. 2: Migrations in Norway as a function of various characteristics

The second example concerns migrations between Norwegian regions observed through the population register, whose data were centralized and computerized in 1964.[2] The complete residential history of each of these persons is known from 1964, but the Norwegian statistical authorities' concern to preserve individual anonymity led us to work at the regional level, Norway being divided into 19 regions (Baccaïni and Courgeau, 1996). We excluded persons having made at least one stay outside Norway.

Our observation covers all men born in 1948, residing in Norway in 1991 and never having migrated abroad, i.e., a total of 28,462 persons. We examine the inter-regional migrations performed in a two-year period after the 1970 census. On the basis of aggregate data, table I.3 reports the number of persons with selected characteristics and the population of the cohort studied for each of the 19 regions. We shall begin by examining the probabilities of migration for farmers compared with other members of the population, without considering individuals in other occupations (column 4) or married individuals as such (column 5).

[2] We would like to thank Statistics Norway for permission to access the files generated by Kjetil Sørlie and Øjsten Kravdal from these data and from population censuses.

TABLE I.3. - MIGRANTS, FARMERS, PEOPLE IN OTHER OCCUPATIONS, MARRIED PERSONS, AND POPULATION OF 19 NORWEGIAN REGIONS

Region	Migrants	Farmers	Other occupations	Married	Population
Ostfold	163	73	1,155	462	1,556
Akerhus	356	58	1,565	595	2,059
Oslo	536	21	2,428	755	3,282
Hedmark	203	103	869	274	1,248
Oppland	231	104	814	285	1,223
Buskerud	168	40	946	314	1,321
Vestfold	164	35	863	328	1,266
Telemark	147	33	778	346	1,150
Aust-Agder	101	30	465	150	632
Vest-Agder	138	38	695	255	937
Rogaland	187	126	1,345	550	1,939
Hordaland and Bergen	260	98	2,073	742	2,862
Sogn Og Fjordane	148	93	489	134	775
More Og Romsdal	247	154	1,163	447	1,840
Sor-Trondelag	204	92	1,144	411	1,656
Nord-Trondelag	205	129	579	243	946
Nordland	373	191	1,225	397	1,942
Troms	210	132	697	207	1,128
Finnmark	133	90	460	140	707

Source: Norwegian population register

From these data, we can compute a linear regression—weighted by regional population—between migration rates $y(r)$ and the proportions of farmers $x(r)$ in each region. Such a weighting, impossible with Durkheim's data, can be performed here: it enables us to rank each region consistently with its population size. In theoretical terms, this is far more satisfactory (Robinson, 1950), and allows clearer comparisons with individual estimates for the total population, which we shall examine later. As a rule, however, the numerical differences between the weighted and unweighted estimates are negligible. This weighted estimate supplies the line in figure I.2 and yields the parameter estimates given in table I.4.

The adjustment quality measured by the adjusted R^2 test gives a value of 0.191, which is fully significant. The estimated probabilities show a positive, significant link between the migration rate and the proportion of farmers in each region. The variance share explained is 0.24 and differs significantly from zero at the 5% level. We can therefore conclude that Norwegian farmers are almost six times as likely to migrate than other occupations, if the model's hypotheses are verified. This result may seem surprising given the financial and personal cost to farmers of moving to another region. We shall also see that some of the hypotheses underlying this estimate may be incorrect.

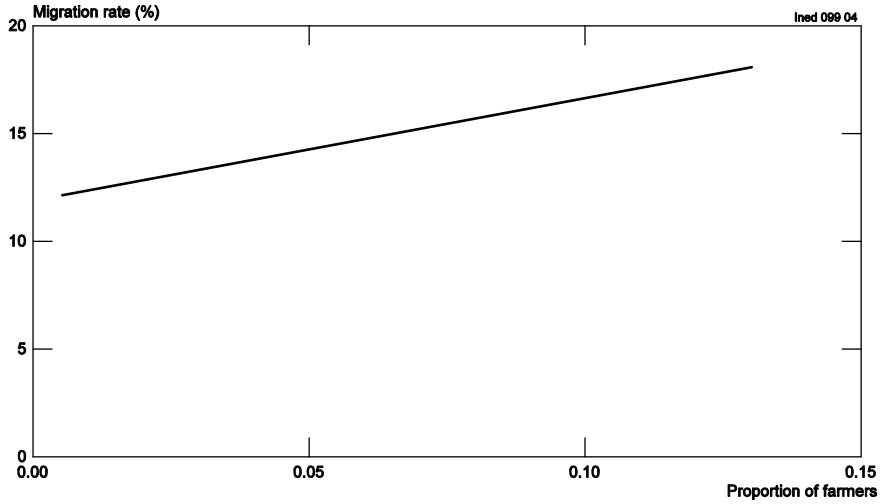


Figure I.2. - Migration rate by proportion of farmers (Norway)

TABLE I.4. - ESTIMATED PARAMETERS FOR REGRESSION MODEL GIVING MIGRATION RATES BY PROPORTIONS OF FARMERS AND PERSONS IN OTHER OCCUPATIONS OR WITHOUT OCCUPATION

Probabilities of migration	Regression on percentage of farmers	
	Estimated value	Standard deviation
Other occupations or without occupation	0.119	0.014
Farmers	0.597	0.197
Adjusted R^2	0.191	

We can perform a similar analysis on persons in non-farming occupations ("other occupations" columns of table I.2). Their probability of migrating is 0.047 and does not differ significantly from zero. Its standard deviation is 0.132.

It should be noted here that the probabilities estimated by linear regression methods can produce rough estimates. In particular, this may occur when the proportions of individuals possessing the characteristic studied cover a small variation interval. By contrast, if the proportions effectively cover the entire interval, the risk will be lessened. To take the first example, the percentages of Protestants vary from 30% to 95%, covering almost the entire interval from 0% to 100%. In this case, we may assume that the estimations of probabilities of suicide by Protestants and other religions are not too biased. But in the second example, the percentages of farmers vary more modestly between 0.6% and 13.6%. While the probabilities of migration for non-farmers are presumably estimated correctly, the estimates for farmers are rougher, even if the model's other underlying hypotheses are properly verified: the extrapolation of the regression results to 100% of farmers produces a 95% confidence interval between 0.21 and 0.98 for the probability of migrating. The same goes for other occupations, whose percentages vary from 61.2% to 76.0%: in this case, the estimated probabilities of the two groups are approximative. Note that in all these examples we are working on populations observed exhaustively, not on samples. Later, we shall consider cases where these estimates are negative or greater than unity and hence non-significant.

Thus far, we have explored a simple case involving the observation of a single characteristic of the population. Let us now see what happens when the observation is more detailed.

Observing several characteristics

The introduction of several characteristics (social, political, economic, etc.) into the same model will raise new problems and greatly complicate the estimation of the probabilities of experiencing the event.

Let us take the case where we observe two characteristics: from there, we can easily generalize the results obtained to any given number of characteristics. Let us continue to assume the homogeneity of the population observed as regards the probabilities of experiencing the event studied when individuals have different characteristics—a precondition for such an aggregate estimation. The probability is thus independent of the area of residence

and is determined only by the two characteristics observed. We want to show that, in this case, we need to estimate four probabilities from the observations.

Specifically, we need to distinguish the probabilities of experiencing the event in the two sub-populations possessing only one of the characteristics, p_{10} and p_{02}, in the sub-population possessing both characteristics, p_{12}, and in the rest of the population, p_{00}. Note that, as a rule, the four probabilities are independent of one another.

To estimate them, we must observe in each region r the corresponding sub-populations: $X_{10}(r)$, $X_{02}(r)$, $X_{12}(r)$, and $X_{00}(r)$. Let us initially assume that we have done so, even though this new hypothesis is seldom verified. We can then write the mathematical expectation of the number of events observed:

$$E[Y(r)] = X_{10}(r)p_{10} + X_{02}(r)p_{02} + X_{12}(r)p_{12} + X_{00}(r)p_{00} = \\ X_{10}(r)(p_{10} - p_{00}) + X_{02}(r)(p_{02} - p_{00}) + X_{12}(r)(p_{12} - p_{00}) + X(r)p_{00} \quad \text{(I.3)}$$

where $X(r)$ is the total population of the region r. As before, we obtain the following relationship:

$$y(r) = x_{10}(r)(p_{10} - p_{00}) + x_{02}(r)(p_{02} - p_{00}) + x_{12}(r)(p_{12} - p_{00}) + p_{00} + \varepsilon_r \quad \text{(I.4)}$$

Again, we can estimate the different probabilities with the aid of the sub-populations observed.

As previously, we have no guarantee that the estimated probabilities effectively lie in the expected intervals between 0 and 1. Moreover, insofar as non-negligible correlations may exist between the x_{ij} values, the parameters estimated by this regression method are no longer interpretable, as we shall see in the discussion of methodological issues raised by this approach.

Lastly, if we do not know the sub-population $X_{12}(r)$, we can no longer estimate the three unknown probabilities. However, there is a special case where the characteristics divide the population into unrelated sub-groups. If so, no individual will possess two or more characteristics simultaneously, so $p_{12} = 0$. We can thus estimate the mathematical expectation of the total number of events experienced in the region $y(r)$:

$$E[Y(r)] = X_{1.}(r)p_{10} + X_{.2}(r)p_{02} + [X(r) - X_{1.}(r) - X_{.2}(r)]p_{00} \quad \text{(I.5)}$$

where $X_{1.}(r)$ and $X_{.2}(r)$ are the sub-populations each possessing one of the characteristics.

This occurs, for example, if we break down the population studied by broad occupational category: at the time of observation, each individual will have only one status. Under this condition, it will be possible to estimate the following model:

$$y(r) = x_{1.}(r)(p_{10} - p_{00}) + x_{.2}(r)(p_{02} - p_{00}) + p_{00} + \varepsilon_r \quad \text{(I.6)}$$

where $x_{1.}(r)$, $x_{.2}(r)$, and $y(r)$ are the proportions observed for the various groups in each region: again, this is a linear-regression model. It will provide estimates of the probabilities, with the same misestimation risks as before.

For a more detailed presentation of the estimates with simulations showing the impact of the non-verification of certain hypotheses on the estimated probabilities, see Courgeau (1999b).

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

To continue the second example, let us distinguish farmers from other occupations and economically inactive persons. The regional numbers are given in table I.3 above. With the estimated parameters of this regression, we can compute the probabilities of migrating for the economically inactive, farmers, and other occupations: they are reported in table I.5.

TABLE I.5. - ESTIMATED PARAMETERS FOR REGRESSION MODEL GIVING MIGRATION RATES AS A FUNCTION OF PERCENTAGES OF FARMERS, OTHER OCCUPATIONS, AND PERSONS WITH NO STATED OCCUPATION

Probabilities of migration	Estimated value	Standard deviation
Farmers	0.597	0.226
Other occupations	0.119	0.181
Occupation not stated	0.120	0.247
Adjusted R^2	0.140	

The probability of migrating for farmers is identical to the one estimated above, whereas the probability of migrating for other occupations and the economically inactive is practically identical at 0.12. There is a difference between the probability of migrating for other occupations computed earlier (0.05) and the one estimated here, but recall that this term consistently displays a very high uncertainty and cannot be regarded as significantly different from the estimated value given previously. This situation is identical to those described earlier, where the intervals in which the characteristics vary do not in any way cover all possible intervals.

Furthermore, for a multiple linear regression of this kind to be correct, the two characteristics must not be strongly correlated. Now we find a very strong negative correlation of -0.85 between the proportions of farmers and those of other occupations. This complicates the interpretation of the estimated probabilities of migrating for non-farming occupations.

The previous formulas can easily be generalized to cases involving a larger number of characteristics, but the interpretation problems are even worse than with two characteristics. We shall therefore refrain from examining them further here, but later on we shall give an example of an application to a situation with multiple characteristics.

Current extensions of the approach

The observation of a population at successive moments in its history led to the theory of demographic transition. Landry (1909) offered a partial formulation, taken up by Notestein in 1945. In this section, we describe its broad outline, referring the reader interested in a fuller presentation to Burch (1999).

The theory aims to supply a universal framework to explain long-term changes in the population of different countries, and to relate those changes to the concurrent demographic, economic, sociological, ecological, psychological, and other changes in the world. It is thus effectively an extension of cross-sectional analysis, for it seeks to explain developments in a given period by equally cross-sectional characteristics. But it is more general as it incorporates observations specific to different social sciences such as economics, human geography, and psychology.

The demographic indicators used will be period indicators such as mortality rate, birth rate, immigration rate, international emigration rate, infant mortality rate, life expectancy, total fertility rate, and so on. They usually concern regions or nations, but can cover larger sets such as the developed countries as a whole. The explanatory factors seek to characterize different moments of radical breaks in the country's history: agricultural revolution, industrial revolution, and dissemination of new standards or new inventions such as contraceptive methods and cultural changes.

Demographic transition theory was developed from an initial framework showing that industrialization and health progress were key drivers in the transition (Davis, 1945). Numerous observations show that the changes caused by these two general phenomena initially produce a decline in mortality, which, after a certain lag, leads to a decrease in fertility. This is caused by the increase in family size due to the survival of a greater number of children and to the higher cost of providing for these families. However, this explanation is not always valid: in some countries or regions, the decline in mortality followed rather than preceded the decrease in fertility (Coale, 1973), contradicting the universality of the sequence. Hence the need to observe in greater detail the patterns in developed countries as well as in developing countries.

A more geographically-oriented approach is therefore required. Zelinsky (1971) introduced the transition into geographic space. He showed that, by adding urbanization and spatial-diffusion phenomena to those mentioned above, one could better understand the spatial-mobility transition in parallel with the demographic transition. This spatial approach was fleshed out by Cleland (1985), who attributes the lags in the start of transition in different countries to the dissemination of information and the introduction of new social norms for birth control.

Economists have also contributed their views on this transition. Becker (1960) and Schultz (1973) emphasize three characteristics that provide a measure of couples' choices: (1) relative costs of children compared with other goods; (2) couples' incomes; (3) couples' preferences between raising children and other forms of consumption. Easterlin (1961) speculates that women in large cohorts tend to have fewer children than those in smaller ones, causing what are assumed to be cycles in period fertility. The verification of this theory remains unsatisfactory. In a fuller formalization of his approach (Easterlin and Crimmins, 1985), Easterlin eventually developed a new economic model, incorporating sociological characteristics: the number of children that parents would have in the absence of voluntary birth control ("supply of children"), the number of live children that they would like to have ("demand of children"), and fertility-regulation costs, which are psychic, social, and financial. However, the simulations of these models on observed data do not yield fully convincing results.

To the economic arguments, Lesthaeghe (1983) adds a shift in psychological values toward individualism and the quest for self-fulfillment, which derive from the secularization of society and its growing wealth. Caldwell (1982) introduces the feeling that causes the

replacement of parents by children as the recipients of economic benefits ("emotional nucleation of the family") and produces an intra-family reversal of wealth flows.

More generally, these theories lead to a systemic approach to population change. This approach holds that the economic, social, cultural, psychological, political, and other components of a given society do not vary independently but on the contrary evolve as a whole. The system's complexity leads to modeling by means of macro-simulation or micro-simulation (Van Imhoff and Post, 1997, 1998; Burch, 2002). Unfortunately, such models often produce a simplification of the links between the system's elements, which can quickly result in very different changes from those ultimately observed.

An interesting solution to these problems was offered by Bonneuil (1997, 1998), who uses the theory of viability (Aubin, 1991). As in the model described earlier, he argues that a group's evolution can be represented by a small number of period characteristics. The group is subjected to a set of constraints and, in consequence, its survival paths are not determined at random. They will depend on several controls on which the group can act. This theory enables Bonneuil to estimate the viable paths among all the possible alternatives, and to identify the periods of stagnation and the leaps between norms. Applied to the changes in Swedish fertility since 1930 (Bonneuil, 1994), the model proves more effective than Easterlin's hypothesis in grasping the abrupt shifts due to World War II.

As we can see, this approach is still very much alive, insofar as it treats the society studied as a whole whose overall behaviors are dictated by a set of general rules. There is no need to analyze individual behaviors, for

> "society is not a mere sum of individuals, but the system formed by their association represents a specific reality with its distinct characteristics" (Durkheim, 1895, p. 102).

Let us see how we can interpret this approach more precisely.

II. Underlying paradigm

Although past authors have never clearly articulated this frame of reasoning—also called a paradigm—we can deduce it from the presentation above. The paradigm defines the norm of what is a legitimate activity inside the scientific domain that it governs (Kuhn, 1972). A paradigm is therefore intrinsically resistant to a precise definition, but we can outline it by means of sufficiently general principles. For instance, we can say that social facts exist independently of the persons who experience them. They are explained by various economic, political, religious, social, and other characteristics of society: this defines a form of causality originating in society itself and not in the individual, a causality whose effects are felt on an entire population. The data-collection method follows logically: it consists in the fullest possible recording, at regular intervals, of the population, its characteristics, and those of the places where it lives. This approach is thus indeed a form of *holism*, in that it explains the evolution of a society in terms of its overall goals, without bringing individual will into play.

We see the establishment of this paradigm during the period in which political arithmetic gained ground, with the aid of data from registers viewed from a cross-sectional perspective; from the early nineteenth century onward, the paradigm asserted itself with the institutionalization of national statistical offices and the introduction of censuses. These provide comprehensive snapshots of the population and record numerous characteristics of both individuals and households. Together with vital statistics—also examined from a cross-sectional point of view—censuses allow an exhaustive analysis of mortality, fertility, and

nuptiality at a given moment. When population registers are also available, or when censuses ask people about their place of residence at an earlier point in time, demographers can examine internal migrations inside the country in the same manner.

Durkheim's analytical method is perfectly suited to the task. Durkheim himself recognized that statistics offered the means to isolate social facts:

> "They are, indeed, represented, not without accuracy, by the birth rate, the marriage rate, and the suicide rate, i.e., by the number obtained by dividing the annual mean total of marriages, births, and self-inflicted deaths by that of men old enough to marry, procreate, and commit suicide." (1895, pp. 9-10)

Admittedly, we can produce period indices more sophisticated than this overly simple rate to properly track the phenomena studied, but the principle remains the same.

Such indices allow the use of regression methods, which will link them not only to one another but—more importantly—to various aggregate characteristics of the areas where individuals live. We have already noted some of the issues raised by these methods. Let us now take a closer look, therefore, at the methodological problems posed by this aggregated cross-sectional approach.

III. Methodological issues

We shall distinguish between different types of problems posed by the use of this approach. We begin by the questions on the interpretation of synthetic indices, used in the descriptive approach. Next, we look at the difficulties in regression-model estimation caused by the approach. In conclusion, we discuss the ecological fallacy induced by an erroneous interpretation of the results of these regressions.

Synthetic indices that are hard to interpret

The first questions on the cross-sectional approach were raised by the use of synthetic indices build from period tables. The indices were compiled with the aim of answering some wholly legitimate questions. Sometimes, however, they yielded results that were hard to interpret and even in logical contradiction with the phenomenon they were supposed to measure. For example, when we want to estimate a probability of survival at any given age, we can always combine the complements to unity of the period probabilities of dying, from birth up to the age examined. But we can easily see that these probabilities bear no possible relationship to any actual cohort, as they measure the effect of period conditions on mortality—such as an epidemic or a severe winter—on hypothetical cohorts. The comparison of such an estimation between different populations or sub-populations of the same country examined in the same period is therefore not as self-evident as appears at first sight. In particular, one of the requirements is that

> "the various cohorts should not reach the start of the year studied with specific histories that will largely determine their mortality during the year" (Pressat, 1966, p. 137).

As we can see, the aim here is to obtain results concerning cohorts by means of cross-sectional analysis, whereas in fact we are working on a hypothetical cohort unrelated to any actual cohort.

Even worse difficulties emerge in the study of other phenomena, such as fertility or nuptiality, which involve periods of delay followed by periods of recovery after an economic crisis or a war. For instance, as Henry explained (1966),

> "during a recovery period, behavior is influenced by the previous delay; assigning to a hypothetical cohort a series of indices observed in a recovery period is thus tantamount to postulating the existence of a cohort that, from one end of its life course to the other, would strive to make up for time it had never lost" (p. 468).

This explains why age-specific overall first-marriage probabilities, which measure the intensity of nuptiality and should always be below unity in an actual cohort, can assume values far greater than unity in a hypothetical cohort. For example, they exceeded 1.5 in France in 1946, immediately after World War II.

Likewise, the assumption that behaviors are influenced only by economic, political, social, and other conditions of the period raises an ever greater number of questions. The demographic problems arising from war are not period problems but concern cohorts that have experienced them longest. More generally, it is important to emphasize the effect of basic factors, which are far more closely linked to cohorts who, as Ryder (1965) notes, "share a common historical location" and have undergone the same experiences at the same ages. Period factors are actually experienced in very different stages of life by each cohort and they can have equally different consequences. An economic crisis experienced by young people can offer them an opportunity to hold different jobs, between which they later will be able to choose in a well-informed manner; by contrast, the same experience for older persons can drag them into successive unemployment spells from which they will be unable to extricate themselves.

Regressions that are hard to interpret

The next source of problems is the interpretation and estimation of the probabilities of occurrence of the event studied, on the basis of the characteristics of the population examined. We have already noted these difficulties; we must now analyze them in greater detail.

The first type of error concerns the case where we observe only one characteristic as well as the case where we observe any number of characteristics. It is linked to the fact that the parameters estimated with the regression model are constrained by the initial hypotheses. The parameters are supposed to allow the estimation of the probabilities of the event's occurrence in homogeneous sub-populations in the entire territory studied: these probabilities should therefore lie in the interval [0,1]. However, in regression models, nothing requires estimated coefficients to obey these constraints (Bry, 1996) and, as the following example shows, the constraints can be largely overcome. But this indicates that some of the aggregated hypotheses of the cross-sectional model are not verified.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

Let us see the effect of the fact of being married on the probability of migrating between Norwegian regions. The number of married persons in each region is also given in table I.3. From these figures, we can estimate the probabilities of migrating for married persons and the rest of the population. The results are reported in table I.6.

TABLE I.6. - PARAMETERS ESTIMATED FOR THE REGRESSION MODEL GIVING MIGRATION RATES ACCORDING TO WHETHER PERSONS ARE MARRIED OR NOT

Probabilities of migrating	Estimated value	Standard deviation
Unmarried	0.305	0.054
Married	-0.336	0.163
Adjusted R^2	0.304	

We see that if the estimated probability of migrating of the unmarried, 0.305, does lie within the interval [0,1], that of the married is a negative -0.336, which is an absurd result. In this case, therefore, the aggregate cross-sectional model is not applicable, as its hypotheses produce impossible results.

The second type of error arises when we consider the effect of two or more characteristics: it is linked to the fact that correlations may exist between characteristics, even when these are linearly independent. In this case, when we perform a regression on a single characteristic, the probability of experiencing the event, for individuals who have done so, can be very different from the value found when we perform the same regression on this characteristic and on another, correlated characteristic (Bry, 1996). For this effect not to appear, all the explanatory variables examined must be totally uncorrelated with one another. As the following example shows, that is rarely the case in the social sciences.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

Let us now examine several other characteristics besides the fact of being a farmer or having another occupation. For instance, we have distinguished between groups of married and unmarried persons, with or without children, whose probabilities of migrating may differ. We can easily calculate the correlations between these characteristics (table I.7).

TABLE I.7. - CORRELATIONS BETWEEN DIFFERENT CHARACTERISTICS THAT MAY INFLUENCE NORWEGIANS' MIGRATION

	Farmer	No stated occupation	Married with child(ren)	Married with no children	Unmarried with child(ren)
Farmer	1.0000				
No stated occupation	0.0117	1.0000			
Married with child(ren)	-0.3422	-0.0141	1.0000		
Married with no children	-0.7486	-0.1671	0.4163	1.0000	
Unmarried with child(ren)	0.6143	0.0642	-0.2902	-0.6394	1.0000
Over 12 years of education	-0.7951	0.0234	0.1059	0.6551	-0.5840

As we can see, many characteristics are strongly correlated with one another, so their respective effects on the probability of migrating are usually indistinguishable. We performed the regression using all characteristics simultaneously, but we have not reported the results here, as they yield only a single significant parameter: the one concerning farmers. Moreover, some of the estimated probabilities generate results outside the interval [0,1]. When we try to introduce the other characteristics incrementally, we find no other significant effect.

The ecological fallacy

If the use of aggregate data is legitimate under the proposed paradigm, the use of individual data may call the paradigm into question. Thanks to new methods such as logistical regressions, we can now estimate models based on individual characteristics. And this estimation may not yield results equivalent to those formerly obtained with aggregate data.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

For instance, the positive link between the percentage of farmers and the percentage of migrants—found above with Norwegian data—shows only that the highest probability of migrating is tied to a high proportion of farmers. We cannot say whether the group with the largest proportion of migrants in each region consists of farmers or non-farmers. The only evidence for the argument that occupation plays a dominant role in migration comes from the hypothesis that social facts exist independently of the individuals who experience them: if so, farmers are the group most likely to migrate in the different regions. But it is entirely possible that the presence of a high percentage of farmers will restrict job offers in other occupations, forcing non-farmers to migrate in search of better jobs—whereas the farmers themselves would have no reason to migrate more than other groups outside their region. Under this scenario,

the interpretation offered in the present chapter no longer applies, and we need to look elsewhere for a better interpretation of the data observed. Later, we shall see that only a multilevel model can address the problem.

A problem of this kind was identified more than 50 years ago (Robinson, 1950). At the time, many researchers were working on data aggregated by region, geographic division, etc., to minimize operating costs or quite simply because the data were most often available only in aggregated form. They believed the results obtained could be directly interpreted in terms of individual behavior. Robinson clearly showed that the correlations between two characteristics measured in binary mode on individuals, or by proportions applied to different geographic segmentations, generally diverged. For instance, in 1930, the correlation between the proportion of the population born abroad and the percentage of illiterates in the United States was -0.526 at the regional level, whereas the correlation between birth abroad and illiteracy for an individual was 0.118. Here, the results at the individual and regional levels were totally contradictory. Robinson concluded his article on a highly categorical note: one cannot use an ecological correlation, measured at the aggregate level, as a substitute for an individual correlation. Some authors have extended these results to analyses using linear-regression methods (Duncan *et al*., 1961; Alker, 1969).

In this case we are confronted with what is known as the *ecological fallacy*, meaning that aggregate data, as a rule, cannot be used to study individual behaviors. The only instance where this is possible is when the probability of experiencing the event is independent of the area studied and when the area's population is large enough to cancel out any random differences that may appear.

The approaches examined in the following chapters will seek to resolve these various difficulties. As we shall see, their resolution was accomplished in several stages.

CHAPTER II

INTRODUCTION OF SENIORITY IN THE GROUP

The instantaneous vision of period analysis deprives human life of all its density, as the analysis applies to a given moment and supposes that demographic phenomena are determined by the characteristics that the population studied displays immediately prior to the instant in which the phenomena occur. In this approach, events occurring in a relatively distant past or characteristics encountered in earlier periods cannot influence events in the period studied. For instance, an economic crisis or a war can affect various demographic behaviors when it occurs, but once it is over it can no longer influence later events. While such an event will obviously entail subsequent recovery spells, period analysis is incapable of capturing them.

As noted in chapter I, there exists an alternative vision of time that could resolve some of the problems encountered with period analysis. By giving precedence to historical time, the period approach emphasizes the instantaneous; by consolidating results for the different cohorts observed in that moment, it yields synthetic (i.e., overall) indices for a hypothetical cohort that, in some cases, raise serious objections. This artificial construct takes the timing changes in the real cohorts and turns them into apparent changes in the mean number of events per head in the hypothetical cohort. In contrast, by emphasizing a time linked to the seniority of individuals in a given status, the cohort approach stresses duration and allows a better separation between the fundamental and the transient (Henry, 1959).

Although some earlier voices had suggested that period analysis was perhaps not the best approach (Delaporte, 1941), it is essentially from the end of World War II that demographers—most of them French (Louis Henry and Roland Pressat, for example)—introduced cohort analysis and showed its ability to solve some of the problems inherent in the period approach.

As before, we begin by examining how this analysis was put into practice and how it operates; next, we identify its basic principles, which we can then discuss specifically.

I. - Introduction of time lived by the generation or cohort

To implement the approach, we need to define: (1) the population(s) to be studied, (2) a temporality common to all members of this population, and (3) the relationships between the phenomena examined.

Definition of populations studied and of the temporality used

Generational analysis initially works on *generations*, born in a given year or period. It tracks them throughout their lives, from birth, to see how demographic phenomena unfold: marriages, childbirths, migrations, occupational changes, and, lastly, the death of the individual. The time unit studied is therefore identical for all cohort members: it consists of their *age* at the occurrence of each event. It may, in fact, be difficult to observe and measure this age accurately, but we shall not discuss the issue in detail here because it lies outside our chosen subject. We should also note that the purpose of generational analysis is not to examine individual event histories such as these, but rather to analyze the collective history of a generation or group of generations. In other words, we are still attempting to analyze *aggregate data*: the groups considered will no longer be a set of sub-populations observed at a given moment, as in period analysis, but a set of sub-populations issued from a given generation and observed at each age.

For a better tracking of changes in certain phenomena that are very closely linked to others (a mother's legitimate childbirths imply her earlier marriage), it is clear that the person's age at the occurrence of the various events may not be the best indicator to follow. If we are studying married women, for example, it seems preferable to look at the births of their legitimate children starting from marriage, and—for births of order higher than 1—from the earlier birth. We can thus define a *cohort* more generally as the set of persons who have entered a given population category in a given period, such as a calendar year. Their entry may be due to a demographic event but, equally well, to any other event in the broad sense: a change in attending physician for a patient; a change of class, even without a special exam, for a student, etc. The time considered will be the time elapsed since this initial event: we shall call this the *seniority* in the group.

Being able to choose from a large number of initial events and, therefore, of different temporalities does gives us high flexibility in our investigation, but it does not solve the problem of selecting the most satisfactory time unit to track in the study of a given phenomenon. Later on, we shall explore possible solutions, but we have to admit that classic cohort analysis does not allow us to do so.

Relationships between phenomena and population homogeneity

Whatever the temporality chosen, the phenomena studied will seem inextricably interlinked: a death can prevent a marriage, a change of occupation can change the probabilities of migration, and so on. In theory, this tangle of facts is not easy to unravel, although we shall see later that event-history analysis enables us to do it better. The difficulty is compounded by the fact that the official statistics available to the promoters of cohort analysis (mainly vital statistics, more seldom population registers) were very succinct. For example, in the case of a death that could prevent a marriage, we have no information on the potential marriage, even if it was imminent. Changes of occupation are recorded neither in vital statistics

nor in population registers. We thus need to formulate simplifying hypotheses on the relationships between demographic phenomena. One solution is to isolate a single phenomenon—the one we have chosen to study—and to regard the other phenomena involved in the analysis as *disturbing phenomena*.

This interdependence between phenomena will be compounded by the diversity of the characteristics of the members of the population, who have no reason to be equally at risk of experiencing a demographic phenomenon. For instance, for a given life span, a senior manager and a manual laborer clearly have very different probabilities of dying: their ratio is 1:2.8 in France, for men aged 35-60 in 1975-1980 (Desplanques, 1984). Likewise, as period analyses have shown (Roussel, 1971), nuptiality varies widely according to occupation: in the 1968 census, the percentage of unmarried at age 50 peaks at 45% for farm laborers; it runs as high as 38% among farmers, and is lowest for senior managers at 4%. These major differences, observed in cross-sectional analysis, are hard to transpose to cohort analysis, as we would need to track individual occupations over time. We cannot do this with vital statistics, which do not record all the occupational changes in a population. Again, we shall see later that the only way to identify these effects is to conduct event-history surveys.

Faced with this complexity and frequent lack of essential information, how will the demographer make the analysis possible? We will need to simplify as much as possible the relationships between demographic phenomena and assume that the population's characteristics have little effect on their occurrence.

For example, if we assume that the phenomena are independent, will we not be able to analyze them more simply? This means isolating the phenomenon studied by eliminating the effect of the other phenomena, regarded as disturbing. Likewise, if we can choose groups that are sufficiently homogeneous with regard to the phenomenon studied, will we not be able to study them more simply? Let us see, therefore, how cohort analysis emerged by simplifying situations, making necessary but heroic assumptions, and exploiting these various possibilities.

Implementation of cohort analysis

Cohort analysis rests on the observation in vital statistics of the main demographic events experienced by a cohort. We shall attempt to separate the influence of the phenomenon that we want to study from that of the other, disturbing phenomena, which modify the populations at risk of the event.

First, we need to formulate a hypothesis on the probabilities, for different members of the population examined, of experiencing the selected phenomenon as well as the interfering phenomena. We assume that to each member of the generation or cohort, for each length of stay, there corresponds a set of probabilities of occurrence of the events studied and of disturbing phenomena. These probabilities are identical for all members of the population. We therefore suppose that the cohorts are *homogeneous* in regard to each event.

But this condition is not sufficient to estimate the probabilities of occurrence of the phenomenon studied by simply eliminating the effect of the other phenomena. We must further assume that the persons who first experienced the disturbing phenomena displayed the same behavior toward the phenomenon studied as those who did not experience it. We therefore need to posit independence between the disturbing phenomena and the phenomena studied. For instance, if we want to study the nuptiality of a homogeneous population at a given age, we

must assume that mortality and international migration at that age and at earlier ages are independent of a person's marriage (Henry, 1972).

Once we have stated these two hypotheses, we can estimate, for each seniority, a probability of occurrence of the phenomenon studied (Henry, 1959; Pressat, 1966) from the numbers of individuals observed in the vital statistics. For each age x, if we are working on a generation, or for each duration, if we are working on a cohort, we know the number of individuals at risk at the start of the year, S(x), the number of individuals who have experienced the event studied during the year, M(x), as well as exits from observation, deaths, international migrations, etc., D(x). We can easily deduct the classic probability sought for,[3] if the hypotheses are verified:

$$m(x) = \frac{M(x)}{S(x) - 0.5D(x)} \qquad (2.1)$$

These probabilities would be identical to those we would determine if we had a population not at risk of the disturbing phenomena. We then compute the events of the table that we would observe absent the disturbing phenomena, starting, for example, with 100,000 individuals at risk at the start of the observation of the generation or cohort. Their sum to age 50, for example, will give the intensity of the phenomenon studied before this age, and the time distribution of the probabilities will give its timing or tempo.

The concepts of intensity and timing allow an accurate summary of the occurrence of any demographic phenomenon, independently of others—in other words, in its pure state (Henry, 1972). If we take mortality, for example, its intensity will necessarily equal unity, as every person is mortal, and its timing will enable us to see if infant mortality is very high, to show how mortality varies with age, to assess the effect of economic crises, wars or recovery from these events on the timing, and so on. This example illustrates the value of cohort analysis by comparison with period analysis. We can also work on order-specific fertility and define the intensity and timing of each birth. For births of order higher than 1, it may be useful to adopt a temporality consisting of the interval separating them from births of the order immediately below. We can also work on fertility for all orders combined and calculate completed fertility and age-specific rates to see its timing, and so on. For a more detailed application to demographic phenomena and different methods for measuring them, we refer the reader to classic works on demography (Henry, 1972; Pressat, 1966).

The results of such an analysis, which observes a generation or actual cohort, remove some of the drawbacks we had encountered in period analysis. For instance, the intensity of nuptiality (first marriages) will always be less than or at most equal to unity. We shall also be able to measure the effect of an epidemic or a war on mortality, by comparing—at the ages when these external events occurred—the cohorts that experienced them with other, neighboring cohorts that did not.

We can also use data from successive censuses, but the hypotheses are even more restrictive than for vital statistics. For these data to yield satisfactory results, the continuity hypothesis (Henry, 1966) must be fulfilled, i.e., the fact of having experienced the event studied does not influence the probability of the disturbing phenomena that occur after it instead of before (as in the previous scenario). Considering the major differences in mortality between the unmarried, the widowed, and the divorced versus the married, between socio-occupational categories, and so on, we can hardly assume that this condition will be met for mortality. To take the example

[3] For more details on this estimation, see Henry (1972), p. 78. For repetitive phenomena, such as fertility without distinction by birth order, we similarly compute age-specific or duration-specific rates, expressed as ratios to the mean size of the generation or cohort (Henry, 1972, p. 97).

of French males in 1975, their probability of dying at age 55 was 10.9 per thousand for the married, twice as high for the unmarried at 20.9 per thousand, and nearly three times as high for widowers at 28.2 per thousand (Desplanques, 1984, p. 40). For international migration, the continuity hypothesis is even less verified, as individuals having experienced international migration may have behaved very differently from the rest of the population before that migration. As regards nuptiality among farmers, discussed in chapter I, this condition certainly does not obtain, for the exit from farming must be regarded as a disturbing phenomenon: as demonstrated later, there is no reason why married men and unmarried men should have the same motives for leaving agriculture.

Example no. 3: Order-specific migrations in Norway

To estimate order-specific probabilities of migration in Norway, the population register gives us the dates of occurrence of each migration for all members of the population. We are working here on total internal migrations in Norway by women[4] in the cohort born in 1948, from age 15 years for the first stay or from the age at the earlier migration for later stays (we have eliminated all stays starting and ending in the same year). The Norwegian file used does include all these migrations until the end of the observation in early 1992. By way of example, here is the start of the file, which records successive migrations up to the fifth.

Individual number	Migration number	Year started	Year ended	End of stay observed
1	1	1963	1992	0
2	1	1963	1992	0
3	1	1963	1967	1
3	2	1967	1968	1
3	3	1968	1970	1
3	4	1970	1971	1
3	5	1971	1974	1
4	1	1963	1992	0
5	1	1963	1971	1
5	2	1971	1973	1
5	3	1973	1974	1
5	4	1974	1992	0

TABLE II.1. - START OF FILE RECORDING SUCCESSIVE MIGRATIONS (FIRST FIVE INDIVIDUALS)

When the end of a stay is observed, the code entered in the last column is 1; otherwise, it is 0 and the year recorded is that of the end of the observation. We see that the first person stayed in the same municipality from 1963 to 1992 without migrating, the third person migrated more than five times during the observation period (of which only the first five are recorded here), and the fifth person migrated four times, the final stay ending with an exit from observation.

Assuming a homogeneous population and the absence of a link between migration and exit from observation, we can estimate the series of order-specific probabilities of emigration. For

[4] While differences between male and female migrations are small, we prefer to take the example here of female migrations, whose results are more eloquent.

order-1 migrations, the duration is measured by age from 15 years on; for higher-order migrations, we look at the time elapsed since the previous migration. Table II.2 shows, by way of example, the first ten order-2 probabilities of migration, computed from the population initially at risk, the number of order-2 migrations observed each year, and the number of exits from observation by persons who did not undertake order-2 migration beforehand.

Duration in years	Population at start of period	Order-2 migrations during year	Exits from observation during year	Probability in p. 1,000
1	**21,403**	**6,586**	**29**	**307.92**
2	**14,788**	**3,446**	**25**	**233.22**
3	**11,317**	**2,031**	**37**	**179.76**
4	**9,249**	**1,296**	**45**	**140.47**
5	**7,908**	**794**	**34**	**100.62**
6	**7,080**	**543**	**43**	**76.93**
7	**6,494**	**420**	**40**	**64.87**
8	**6,034**	**302**	**39**	**50.21**
9	**5,693**	**254**	**43**	**44.79**
10	**5,396**	**186**	**52**	**34.64**

TABLE II.2. - CALCULATION OF PROBABILITIES OF ORDER-2 MIGRATION BY LENGTH OF STAY SINCE PREVIOUS MIGRATION (FIRST TEN YEARS)

From similar computations on migration orders up to order 5, we obtain the results plotted in figure II.1.

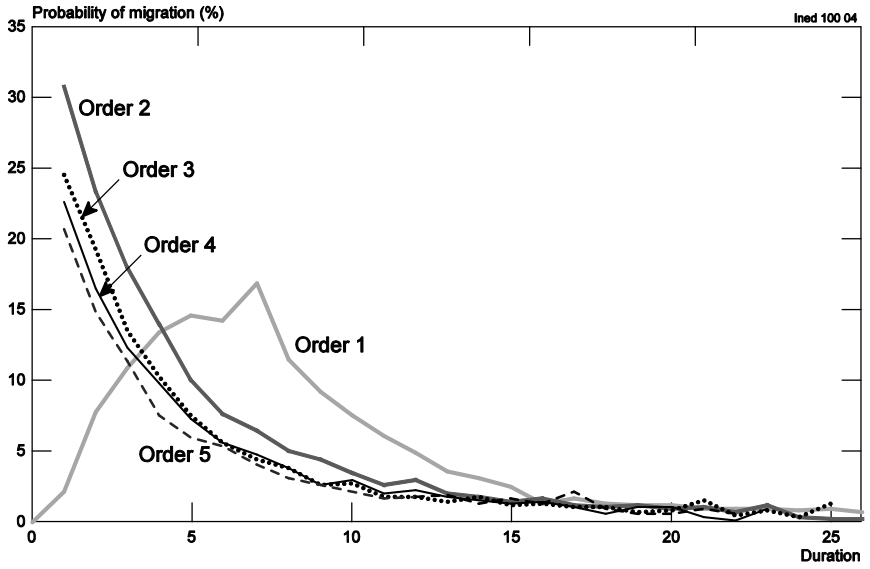


Figure II.1 - Probabilities of migration by order and duration after age 15 for order-1 migration, or by duration since previous migration for other orders

The curve for the first migration is very different from those of higher-order migrations. It rises steadily to age 23, a period marked by entry into the labor market and couple formation, then falls steadily as people settle into working careers and family life. For migrations above order 1, the curves no longer display peaks, but decline steadily with the length of stay. However, we do observe a decrease in the probabilities of new migration with each successive order, at least for lengths of stay of under seven years: beyond, the differences are less clear, notably because of the small size of populations at risk. This may be due to the fact that the higher the order, the greater the proportion of older persons involved in migration. At that point, in addition to length of stay, age at migration may influence the probability. Hence the notion of classifying previous migrations by the migrant's age group: ages 16-20, 21-25, 26-30, and 31 and over. Figure II.2 plots this distinction for migrations of order 2, 3, and 4, displayed as in figure II.1.

We can now see clearly that these probabilities depend on the age at previous migration: the younger the age at which the previous migration occurred, the higher the probabilities, but—for a given age at earlier migration[5]—they virtually cease to depend on the order of the migration considered. These results show that we can model migrations of order higher than 1 with the aid of a small number of parameters, and we try to link the number of migrants measured in a population register with the number of migrants measured by a census question on place of residence at an earlier date (Courgeau, 1973).

[5] The computation of a confidence interval for the probabilities would confirm this observation, but classic cohort analysis never estimates it, claiming that the observation is exhaustive. There is, however, a more basic random factor involved, as individuals follow a stochastic process. Even without sampling, we can estimate the variance of a probability or test whether two probabilities are significantly different or not (Hoem, 1983). The populations observed here, while exhaustive, cannot suffice for long observation spells or for higher orders: the variability observed in figure II.2 shows this clearly.

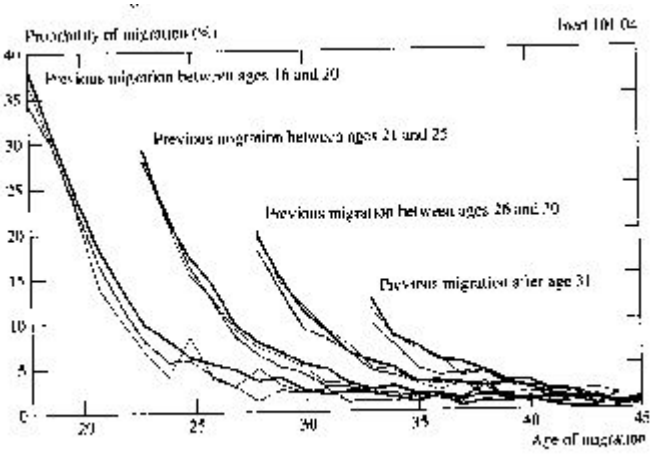


Figure II.2. - Probabilities of order-2, -3, and -4 migration by age at previous migration, and time elapsed since that migration

We noted above that this analysis assumed a homogeneous population, i.e., consisting of identical individuals, whose histories differ only by chance. But actual generations or cohorts consist of individuals who differ in many ways: their social status, earlier history, economic position, etc., are very diverse and generate different behaviors. Period analysis had already demonstrated this clearly. How, then, can we take proper account of this population heterogeneity in cohort analysis? The task seems far less simple than in period analysis, for characteristics will now change with the seniority of individuals in the group. Over time, persons are not only at risk of the event studied, but also experience many other changes in their social, economic, political, and other circumstances: their status will change continuously in all these spheres. Moreover, the sources used to perform these analyses—vital statistics and population registers—generally do not record all these changes in individual status on a continuous basis. We will, for example, be able to determine a person's occupation at the birth of his or her children, but the changes between births go unrecorded. To determine the number of persons in each occupation at risk at each instant, we would need to register the exact date of each occupational change.

The solution offered by differential cohort analysis is to study the occurrence of a given phenomenon in initial groups defined by various characteristics such as occupation, educational attainment, and religion. Although it is hardly achievable with vital statistics, we can take an example to show how it would work. Suppose we want to analyze the nuptiality of a population of individuals initially working in agriculture. We can see that, in addition to mortality and international migrations, we will need to incorporate the exits from agriculture that occur every year. Now while we may assume that the hypothesis of independence between the first two interfering phenomena and nuptiality is broadly verified, the independence between exit from agriculture and nuptiality is far harder to justify. Furthermore, the adjustment to allow for the effects of mortality and international migrations will, as a rule, be very small by comparison with the adjustment needed to take account of exits from agriculture. As we shall see later, the links between nuptiality and exit from agriculture are too important to be ignored.

We can therefore state that differential cohort analysis can hardly solve all the issues raised by heterogeneity, for the latter is generally not an immutable given but will vary over

time. This creates problems due to the heavy dependence between the phenomenon studied and the departure of the sub-population examined, which unquestionably can no longer be viewed as independent. In most cases, therefore, we will have to content ourselves with working on the population in the aggregate without being able to separate it into more homogeneous sub-groups.

Extension to multi-state models

This approach, which analyzed phenomena separately, was extended in the 1960s-70s with models allowing the simultaneous treatment of several demographic phenomena while preserving the condition of their mutual independence. Schoen and Nelson (1974), for example, set up a model analyzing transitions between marriage, divorce, and mortality. Similarly, Rogers's multiregional models (1973) bring regional mortality and fertility simultaneously into play, as well as inter-regional migrations, all these phenomena being studied by age. The models paved the way for multiregional population projections introducing all demographic phenomena. The history of a population or sub-population may be said to follow a process without memory. Mathematicians describe this as a Markov chain, in which "the result of each new trial depends on the one directly preceding it, but is independent of the results of all earlier trials" (Takács, 1964).

II. - Paradigm of the cohort approach

As seen earlier, the goal of the cohort approach is to isolate demographic phenomena in their pure state, so as to rid them of the effect of disturbing phenomena and allow comparisons between countries or periods. We may accordingly define its paradigm by the following postulate: the demographer can study the occurrence of only a single event, during the life of a generation or cohort, in a population "that preserves *all its characteristics and the same characteristics* for as long as the phenomenon manifests itself" (Blayo, 1995, p. 1054). For the analysis to be feasible, the population must also be regarded as *homogeneous* and the interfering phenomena must be *independent* of the phenomenon studied (Henry, 1959).[6] Multi-state models preserve both conditions, applied this time to each of the sub-populations analyzed: the events that such a sub-population may experience occur independently of one another, in conditions specific to the region in which individuals are present.

This analysis implies a denial of all specificity in individual lives; it focuses exclusively on the occurrence of an event, independent of other phenomena, in a population that remains homogeneous over time, being composed of interchangeable units. The cohort approach is thus another example of *holism,* albeit different from the holism that informs period analysis: it denies the existence of individual entities—possessing personal characteristics and observed during their entire lives—in order to confine itself to comparisons between homogeneous groups, observed during their entire lives.

Under this paradigm, we can use data from vital statistics or population registers to extract clear information on the demography of generations or cohorts defined by initial events, such as marriage for the study of legitimate fertility, a birth for the study of a subsequent birth, and so on. We have also noted the possible use of data from successive censuses, but the

[6] Our analysis does not distinguish between renewable and non-renewable events (Henry, 1966). This issue, while important for the analysis, does not require discussion within the scope of our work.

hypotheses are even more restrictive than for vital statistics, as the *continuity* condition must be met for these data to yield satisfactory results.

The behaviors that seem so heavily influenced by current events in period analysis exhibit far greater regularity at the cohort level. The curve of long-term variations in French fertility, which is very erratic when plotted using period analysis, looks far simpler in a cohort analysis. It displays a steady pattern of change from one cohort to the next. From a starting point of more than seven children per woman before 1760, it declines continuously until the cohorts born in 1896, reaching a value of less than two children; it then rises again until the cohorts born in 1930, with 2.6 children, followed by a steady ebb to 2.1 children for the cohort born in 1948; in the recent period, it bottoms out at less than two children per woman for cohorts born in the 1960s.

For non-renewable phenomena such as mortality, order-specific nuptiality, and order-specific fertility, the paradigm allows us to measure their intensity (frequency of never-married status, parity progression ratio for families with *n* children, etc.) and, for renewable phenomena (such as all-orders fertility) the mean number of events per head. We can also estimate the timing of the phenomena (distribution of ages at first marriage, age-specific fertility rates, and so on). We can thus compare, more clearly than in period analysis, the behaviors of different generations or cohorts in a single country and in different countries. The combination of rates in the period approach had the overall effect of turning differences in timing into differences in mean numbers of events per head: for instance, "in period analysis, a gradual decrease in age at first childbirth, without a change in the mean number of children, translates into a steady rise of total fertility" (Henry, 1966). Obviously, in cohort analysis, the differences in timing and mean number of events per head are clearly highlighted.

The demography textbooks informed by cohort analysis examine each phenomenon in a separate chapter, since it is isolated in its pure state: nuptiality, fertility, mortality, migrations, and so on (Henry, 1972). In these conditions, there is no reason to present the possible interactions between the phenomena, since they are regarded as mutually independent.

III. - Methodological issues

Cohort analysis does provide a solution to some problems raised by period analysis and is well suited to the examination of vital statistics. However, it will face difficulties when applied to more detailed data sources on individual event histories. As we shall see, these difficulties are largely due to its underlying hypotheses.

First, the events that determine the entries and exits of the population studied must be defined very strictly. The phenomenon examined must be independent of (1) disturbing events, such as mortality and emigration, that may prevent certain individuals from experiencing it, and (2) competing or concurrent events, such as cohabitation, which "compete" with the phenomenon (in this example, marriage). If independence is not ensured, an obvious selection bias will remove individuals with specific characteristics from the population at risk and, conversely, will admit other individuals who may alter the group's composition. As a large number of events occur in a short period of a person's life, they can interfere with the phenomenon studied: the cohort analysis described here assumes that the latter is independent of each of the others. This hypothesis is fairly implausible when we are looking at events such as entry into the labor market, starting to live on one's own, and the formation of a partnership or marriage, which are bound to strongly influence one another.

Moreover, as the paradigm used here allows the study of only one event, it precludes the study of exits due to competing events. This prevents an analysis of mortality by cause, unless we assume that the causes are mutually independent and can thus be studied separately. It is obvious that the eradication of a cause of mortality, if it is possible, will change the probabilities of mortality from other causes in a way that is practically impossible to predict as long as the first cause persists. Likewise, the study of the exit from never-married status through cohabitation or marriage is hardly realistic, as we need to assume that the two phenomena are independent. Lastly, it is

> "for the same reason that we must forgo the study in a population to which several events allow access (except when we are dealing with different modes of occurrence of the same event, and when the probabilities of experiencing the events studied and the disturbing events do not depend on the mode of occurrence of the 'entry' event" (Blayo, 1995, p. 1507).

That makes a large number of cases indeed in which the paradigm rules out any possibility of analysis.

But the homogeneous-population hypothesis will raise additional problems. As long as we were working on vital-statistics data, this did not seem an obstacle, as we had no other possibility of verifying the hypothesis. The reason is that the vital-statistics source, whose value lies in its completeness, consequently does not give very detailed information on the population studied. While vital statistics will tell us if a birth is legitimate or not, and will give us the birth order, it will provide scant information on the spouses' life histories, which would allow a more detailed analysis of fertility as a function of multiple characteristics. However, it seems obvious that the population on which we are working is in many respects heterogeneous and that this characteristic will influence the probabilities of experiencing the events open to study.

We have already noted the differences in nuptiality observed in a period approach between farm laborers or farmers and other occupations, and shown that the differential cohort approach offered little scope for analyzing these differences over time. But even if we assume this analysis is feasible, could we have isolated a homogeneous population with it? There is no reason to think so, for, as Henry observes (1959, p. 25):

> "[To] determine with precision the practical impact of the heterogeneity of human groups, we shall need to pursue research in differential demography down to the individual characteristics, physical and psychological; in so doing, we must be mindful of studying both the dispersion and the correlation of demographic indices inside the rather roughly defined groups examined until now."
>
> In a note, he adds: "Given the practical difficulties, we will inevitably wonder whether the problem posed is soluble."

We can define the groups by region of residence, for the probability of marriage for a farmer from the prosperous French region of the Beauce will no doubt differ from those of his colleague in the poorer Ariège, depending on their income, educational attainment, and so on. But in the process we shall be dealing with groups so small that their size will prevent any cohort analysis. Moreover, we shall never be certain of having taken into account all the heterogeneity factors in the population. There will be a residual, unobserved heterogeneity, whose effect on probabilities will be totally unknown—contrary to what happens in event-history analysis, as we shall see later.

Also, the possibility for individuals to move between groups will make the analysis even more difficult. The groups defined in the previous paragraph are not stable over time.

Individuals can change occupations or leave the labor force at any time in their life. They can migrate from one region to another, from the Beauce to the Ariège. Their incomes will change continuously over time, with peaks and troughs. Cohort analysis offers no means to measure these changes and will therefore be incapable of factoring them in, whereas they can modify the probabilities calculated.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

Let us return to our earlier example of Norwegian migration influenced by the fact of being a farmer. If we want to pursue our investigation with the aid of cohort analysis, we see that a person, throughout his or her life, may not only migrate but also change occupations, for being a farmer is not a stable status. In example no. 3 discussed in this chapter, we saw that a longitudinal analysis of successive migrations by individuals is perfectly feasible. Ideally, we should now be able to conduct a similar analysis of occupation changes by persons who began their lives as farmers, and to distinguish between shifts to other occupations and returns to agriculture. Unfortunately the Norwegian population register does not allow a continuous recording of occupations over time. The only source of information on occupations is the census, which takes place every ten years, precluding cohort analysis. But even if occupational changes were recorded continuously, the cohort approach would not enable us to conduct a proper analysis to identify the links between geographic and occupational mobility. Thus, in attempting a cohort analysis of data from a retrospective survey on both geographic and occupational mobility, Guy Pourcher (1966) admits that he cannot combine the two mobilities but only analyze them separately.

The introduction of multi-state models will add more difficulties. When we work, let us say, on regional data, the condition of mutual independence between phenomena seems harder to meet than at the national level. For instance, an individual's probability of dying must immediately become identical to those of the area in which (s)he has just settled. But persons who have lived in regions where, for example, death from alcoholism is very high and who will therefore be at risk of being alcoholics themselves will have little reason to change their behavior immediately if they migrate to more sober areas. Likewise, women migrants from abroad—or, conversely, from an area with lower fertility than the destination area—may not immediately adopt the fertility of the latter. We know that, in France, women migrants from southern countries do not adopt the fertility behavior of Frenchwomen instantly, but will do so after a relatively long spell. Similarly, different regional fertility rates may entail rather lengthy adjustment periods.

The probability of inter-regional migration may, of course, depend on a person's age, but will have to be independent of the length of stay in the departure region, earlier migration stages, and the length of stay in each stage. Accordingly, the probability of returning in a previously occupied region is identical to what it would be if the region had never been occupied. Now many examples show that back migrations are far more frequent than migrations to a third location that differs from earlier places of residence (Courgeau, 1982).

Lastly, these models can only incorporate ratios of inter-regional migration to the initial populations, without being able to factor in the populations of the destination regions. However, we know that inter-regional migration flows depend very often on the arrival

population as well as the departure population. The development of models incorporating both populations leads to solutions that are not linear as in all the preceding cases (Courgeau, 1991). Whereas the earlier Markov model generated outcomes with a stable limit state, these new models lead to totally distinct solutions in the long run. These are cyclical solutions and even, in some cases, "chaotic" solutions.

The evidence presented here suggests that major obstacles stand in the way of a strict application of the cohort-analysis paradigm to demographic issues more complex than the separate analysis of single phenomena. To allow for population heterogeneity, the paradigm requires such detailed breakdowns of the population studied as to invalidate any serious calculation. It also imposes such stringent requirements on the events examined that it precludes entire sectors of demographic analysis such as the analysis of competing events, of interacting events, and of events in a population with entry and exit flows. As we have seen, multi-state analysis, far from lessening these problems, makes them even more intractable.

Given all these difficulties, we may well ask if a change in the underlying hypotheses of the analysis is not needed in order to provide a more solid basis for discussing the facts.

CHAPTER III

ANALYZING INDIVIDUAL DATA

The longitudinal vision enables us to fully take into account the time lived by a generation or cohort, whereas the period vision omits it entirely. But this requires a very heroic hypothesis regarding the homogeneity over time of the population examined. The solution that consists in decomposing the population into sub-groups and using differential cohort-analysis methods is unfortunately of scant use here, as shown in the previous chapter: the sub-populations soon become too small to yield significant results, and the situations become too complex owing to entries and exits that fail to meet the independence criterion needed to conduct a proper cohort analysis.

However, the period vision allows a better demonstration of the effect of population characteristics, at a given moment, on a phenomenon studied. But it requires two new and equally heroic hypotheses: (1) the population exhibiting a given characteristic is just as likely to experience this phenomenon throughout the geographic area; (2) the characteristics preceding those of the period studied have no effect on that probability. As a rule, there is no reason for either condition to be met.

To move ahead, we thus need to formulate more realistic hypotheses and apply them using a different approach from the two previous ones. This new approach may require other data than those provided by censuses or vital statistics.

It was in the early 1980s—i.e., more than thirty years after the introduction of cohort analysis—that the new approach first appeared in demography. It had two main objectives, which, for clarity's sake, we shall examine separately. The first objective was to transcend the holism prevailing in the two earlier approaches by introducing an individualistic vision of demographic behaviors. Demographers would thus be able to see how various characteristics of individuals alter their probabilities of experiencing the events studied. Logistic-regression models would prove essential for the purpose. The second objective was to introduce the time lived by each individual, so as to integrate the events in his or her own life history. The introduction of such time regressions—made possible by new statistical methods—simultaneously imposed a new data-collection method: the event-history survey. By combining the two objectives, we shall obtain the methods of event-history analysis in their most general form.

We can then extract the paradigm underlying this approach and discuss more fully the issues it resolves and those that it leaves unaddressed.

I. - Establishing event-history analysis

The event-history approach was implemented thanks to new statistical methods that allowed the processing of discrete data and the introduction of time into the analysis. We begin by examining how to analyze individual behaviors, seen at a given moment, before discussing time regressions, which enable us to deal with behaviors in all their complexity.

Introducing individual behaviors

First, let us set aside time—which we shall discuss in the next section—to concentrate on the potential effect of individual characteristics on demographic behaviors observed in a generation or cohort, at a given moment of their existence. We shall therefore waive the hypothesis of a homogeneous population to see how to introduce its heterogeneity.

In the period analysis of aggregate data, discussed in chapter I, we linked the probability of experiencing an event to the possession of various characteristics, by working on a certain number of sub-populations. To estimate the desired link under certain hypotheses, all we needed to know were the marginal distributions of the events studied. Now, we shall try to estimate the joint distribution of all the combinations of variables, at the individual level, and so to establish a different link between an event's occurrence and the characteristics of an individual. Naturally this requires access to individual data rather than aggregate data as before, but it does allow a clearer estimation of individual effects. We thus arrive at a logistic-regression analysis, which anticipates event-history analysis.

We begin with the simplest case, in which we seek to explain a behavior measured by a binary variable, y_i, equal to 1 if the individual experiences the event studied and equal to 0 if not. The behavior is explained by an individual characteristic which is also measured by a binary variable, x_i. The data needed to estimate the relationship are given in table III.1.

TABLE III.1. - POPULATION SIZE NEEDED FOR ANALYSIS AT THE INDIVIDUAL LEVEL

Has the characteristic	Has experienced the event	
	yes	**no**
yes	N_{11}	N_{01}
no	N_{10}	N_{00}

With these data, we can estimate the probability that an individual with the characteristic will experience the event:

$$P(y_i = 1 | x_i = 1) = p_1 \qquad \text{by} \qquad \hat{p}_1 = \frac{N_{11}}{N_{11} + N_{01}} \qquad \text{(III.1)}$$

and the probability that an individual without the characteristic will experience it:

$$P(y_i = 1 | x_i = 0) = p_0 \qquad \text{by} \qquad \hat{p}_0 = \frac{N_{10}}{N_{10} + N_{00}} \qquad \text{(III.2)}$$

as well as the variances of both parameters:

$$\text{var}(p_1) = \frac{\hat{p}_1(1 - \hat{p}_1)}{N_{11} + N_{01}} \qquad \text{by} \qquad \text{var}(p_0) = \frac{\hat{p}_0(1 - \hat{p}_0)}{N_{10} + N_{00}} \qquad \text{(III.3)}$$

Note that the variances are estimated on the assumption that since all members of the sub-population possess the characteristic, they will be at equal risk of the event. Likewise, those who do not possess it are equally likely to experience the event, but naturally with a different probability from the one above. These hypotheses seem rather implausible and the variances are, no doubt, sharply underestimated. But we have no information to take the estimation one step further. We discuss this problem at the end of the chapter and we shall see, in the next chapter, that—with more specific information on this unobserved heterogeneity—a multilevel analysis offers a better estimate.

Let us apply the new approach to the example presented in chapter I: suicide in Prussia.

Example no. 1 (continued): Suicide in Prussia

Durkheim (1897, p. 152) gave the results obtained at the individual level for Protestants, Catholics, and Jews living in Prussia in 1890. Although he was not working on data concerning a cohort, it is interesting to compare his results with those obtained on aggregate data for each province. Durkheim showed that the suicide rate per million Protestants was only 240 compared with 277 obtained using aggregate data. Moreover—despite the lack of data on the number of Catholics and Jews in 1890 to reconstitute the equivalent of the suicide rate for the other religions combined, we can say that this rate is at least higher than the rate for Catholics (100 per million) and for Jews (180). Thus the ratio of the probability of committing suicide for Protestants compared with other religions is less than 2.77, versus 7.70 with the aggregate data.

Figure III.1, which gives a graphic comparison between the probabilities of suicide for Catholics and Protestants, would be directly comparable with figure I.1, if Prussia was composed of only Catholics and Protestants. Here, we can clearly see the emergence of the ecological fallacy. Unfortunately, Durkheim never realized these major differences between rates obtained with aggregate data and those obtained with individual data, for he never estimated the rates.

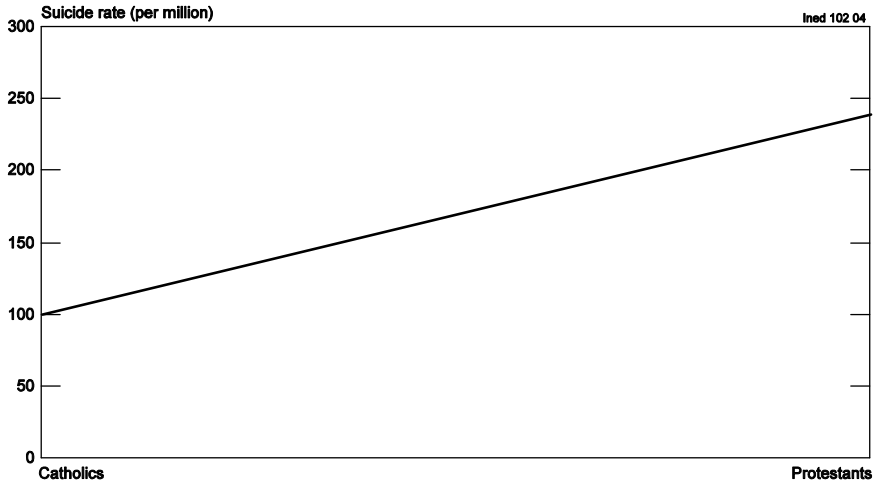


Figure III.1. - Suicide rate for Protestants and Catholics (Prussia 1890)

While the estimation for a single characteristic is simple, when examining a greater number it becomes useful to develop a formal model capable of describing in a satisfactory manner—at least roughly—the relationship between the behavior studied and individual characteristics. Expressing the previous probabilities by a linear model cannot work without constraining its parameters, for the probabilities must range between zero and unity. A simple way to avoid this inconvenience is to use a transformation representing the interval (0,1) on the set of real numbers $(-\infty,+\infty)$. For this purpose, analysts often use the logit function,[7] which in the case of n binary explanatory characteristics[8] is written:

$$P(y_i = 1 \mid x_{i,1},....,x_{i,n}) = (1+\exp-[a_0(1-\sum_{j=1}^{n} x_{i,j}) + \sum_{j=1}^{n} a_j x_{i,j}])^{-1} \qquad \text{(III.4)}$$

In this case, when the individual has the characteristic ($x_{ij} = 1$), all the others being null, his or her probability of experiencing the event is written as a function of the estimated parameters:

$$P(y_i = 1 \mid x_{i,1} = 0,....,x_{i,j} = 1, x_{i,j+1} = 0,...) = (1+\exp-a_j)^{-1} = p_j \quad \text{or} \quad a_j = \log\frac{p_j}{1-p_j} \qquad \text{(III.5)}$$

Gourieroux (1984) gives the methods for estimating this model's $\hat{a}_j$ parameters and their variance; they also allow an estimation of the $\hat{p}_j$ parameters. To estimate the variances of $\hat{p}_j$ we need to use the first terms of a Taylor series expansion of the $a_j = f(p_j)$ functions given above. Indeed, we can write around the parameter's mean value $\hat{p}_j$, assuming that the terms of degree equal to or greater than two are negligible:

$$a_j = f(p_j) \approx f(\hat{p}_j) + (p_{j-}\hat{p}_j) f'(\hat{p}_j) \qquad \text{(III.6)}$$

hence:

[7] We can also use the probit function or the complementary log-log function (McCullagh and Nelder, 1983).
[8] We can also use continuous characteristics, such as income, or polytomous characteristics, such as occupation divided into *n* categories, all combinations being possible. To simplify our presentation, we confine our examination to binary characteristics.

$$\text{var}\, a_j \approx \left[f'(\hat{p}_j)\right]^2 \text{var}\, p_j \qquad \text{(III.7)}$$

From which we deduce:

$$\text{var}\, a_j = \text{var}\left(\log \frac{p_j}{1-p_j}\right) \approx \left(\frac{1}{\hat{p}_j} + \frac{1}{1-\hat{p}_j}\right)^2 \text{var}\, p_j = \frac{1}{\hat{p}_j^2 (1-\hat{p}_j)^2} \text{var}\, p_j \qquad \text{(III.8)}$$

i.e., a standard deviation of:

$$s.d.(a_j) \approx \frac{s.d.(p_j)}{\hat{p}_j (1-\hat{p}_j)} \qquad \text{(III.9)}$$

Most of these estimates are very close to those given by the previous model (as confirmed below with a practical example) and they are therefore open to the same criticisms.

Let us now return to the example of migrations in Norway, discussed in chapter I, which involves observing the members of a given generation during a very short period of time.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

Let us go back to the example of farmer migration. We now have individual data on members of the generation born in 1948 and observed over the two-year period after the 1970 census. We can therefore compute the size of the groups in the theoretical table III.1, which leads to table III.2.

TABLE III.2. - CROSS-TABULATION OF MIGRANTS AND FARMERS FOR AN ANALYSIS AT THE INDIVIDUAL LEVEL

Were farmers in 1970	Migrated in 2 years following	
	yes	no
yes	155	1,485
no	4,019	22,803

We obtain the following estimates of the probabilities of migrating for farmers and non-farmers, as well as their standard deviation:

$\hat{p}_1 = 0.09451$ and $s.d.(\hat{p}_1) = 0.00722$

$\hat{p}_0 = 0.14984$ and $s.d.(\hat{p}_0) = 0.00218$

We can therefore see that, contrary to what we had shown with the aggregated data, the probability of migrating is more than one-third lower for farmers than for other occupations. As already noted, this figure is more consistent with our expectation. The examination of the standard deviations shows that these estimates are significantly different: at the 95% level the confidence intervals are (0.08729, 0.10173) for farmers and (0.14766, 0.15202) for other occupations.

The estimates of the parameters of the corresponding logit model yield the results shown in table III.3.

TABLE III.3. - PARAMETERS ESTIMATED WITH A LOGIT MODEL AND THEIR STANDARD DEVIATION

Individual characteristic	Logit model	
	Estimated parameters	Standard deviation
Farmer	-2.2597	0.0844
Non-farmer	-1.7359	0.0171

We can calculate the quality of the model by testing the maximum likelihood of the estimated model against the model without characteristics. This gives us a value of 15769.66 for χ^2, with one degree of freedom, which is highly significant. We can also calculate the corresponding p estimators:

$$\hat{\hat{p}}_1 = [1 + \exp(-\hat{a}_1)]^{-1} = 0.09452 \quad \text{and} \quad \hat{\hat{p}}_0 = [1 + \exp(-\hat{a}_1)]^{-1} = 0.14983$$

and the estimated standard deviations:

$$s.d.(\hat{\hat{p}}_0) = 0.09452 * 0.90548 * 0.0844 = 0.00722$$

and

$$s.d.(\hat{\hat{p}}_1) = 0.14983 * 0.85017 * 0.0171 = 0.00218$$

Therefore, the estimates obtained with both models are indeed identical, as we noted. We can thus compare figure III.2 below with figure I.2. The individual behavior identified here manifestly contradicts the results obtained with the aid of aggregate data, which also estimate the probabilities of migrating. We shall therefore need to examine in greater detail the hypotheses underlying both approaches, which are not compatible. We can only solve this problem later, once we have better defined the paradigm underlying the event-history approach and have moved toward a multilevel approach.

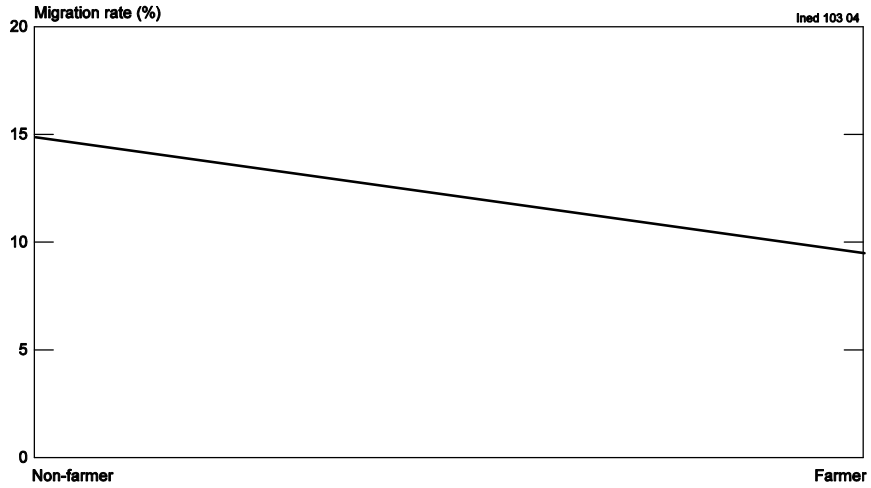


Figure III.2. - Rate of migration of farmers and others occupations in Norway

We can now examine simultaneously the effects of a larger number of characteristics. The logit model allows the estimation of the parameters a_j and p_j corresponding to each characteristic.

TABLE III.4. - ESTIMATION OF LOGIT-MODEL PARAMETERS, PROBABILITIES OF MIGRATING, AND STANDARD DEVIATIONS FOR SELECTED INDIVIDUAL CHARACTERISTICS

Characteristic	Parameter a		Probability of migrating p	
	Estimated value	Standard deviation	Estimated value	Standard deviation
Farmer	-2.3391	0.08503	0.08794	0.00682
No stated occupation	-1.5169	0.03485	0.18069	0.00471
Married with child(ren)	-1.9025	0.04200	0.12983	0.00474
Married without children	-1.2077	0.05115	0.23011	0.00906
Unmarried with child(ren)	-1.6009	0.10201	0.16786	0.01425
More than 12 years' education	-1.5271	0.06843	0.17842	0.01000
None of the characteristics above	-2.0353	0.02553	0.11555	0.00261

We can calculate the quality of the model by testing the maximum likelihood of the estimated model against the model that specifies farmer/non-farmer as the sole characteristic. This gives us a value of 458.01 for χ^2, with five degrees of freedom, which is still highly significant.

We also note that there is little change in farmers' estimated probability of migrating depending on whether it is estimated without applying the other characteristics (0.095) or by applying them simultaneously (0.088). The effects of the other characteristics differ very significantly from one another. For example, married men with child(ren) are almost half as likely to migrate as married men without children, the value for unmarried men with at least one child being in between. When two characteristics—here, being married/unmarried and having / not having children—are rather strongly interdependent, it is very instructive to consider three groups of individuals: those who possess one of the two characteristics but not the other (married without children / unmarried with children) and those who simultaneously possess both (married with children).

It is important to realize that, in this case, contrary to the model using aggregate data, the correlations between characteristics are consistently weak and corroborate these results. Table III.5 shows the correlation coefficients and should be compared with Table I.7.

TABLE III.5. - CORRELATIONS BETWEEN CHARACTERISTICS THAT MAY INFLUENCE MIGRATION BY NORWEGIAN MEN

	Farmers	No stated occupation	Married with child(ren)	Married without children	Unmarried with child(ren)
Farmers	1.0000				
No stated occupation	-0.1421	1.0000			
Married with child(ren)	-0.0383	-0.1332	1.0000		
Married without children	-0.0318	-0.0334	-0.1298	1.0000	
Unmarried with child(ren)	0.0016	-0.0095	-0.0698	-0.0434	1.0000
More than 12 years' education	-0.0442	0.1120	-0.0433	0.0182	-0.0211

This table shows that the correlations between the selected characteristics are very often lower than 0.05 and never reach 0.15. The table is very different from the table of correlations between aggregate characteristics, in which correlations were far higher and seldom below 0.50.

To sum up, this approach analyzes events experienced by individuals during a brief period of time and explains them by the individuals' characteristics at the start of the period. But it has several drawbacks. First, it entails a major loss of information, for it aggregates events taking place in a period—here, three years—and sets aside their exact date of occurrence. Also, it does not take into account the date of settlement in the region of residence inhabited in 1970: this eliminates the effect of the length of stay on the probability of migrating. We cannot observe a time-varying effect of the initial characteristics. Second, the characteristics are set at the time they were measured in 1970 and cannot vary over time. Again, there is no reason why an exit from agriculture should not change a person's probability of migrating. Unfortunately, we cannot incorporate this variation into a logit model, which provides a photograph of the situation of individuals during a very brief interval.

Hence the need to introduce time into such an approach, a step that will transform it into a full-fledged event-history analysis.

Introducing time regressions

Vital-statistics data are of little help in setting aside the second hypothesis of independence between demographic phenomena, for they record only a small number of events, separately and often in a way that makes it difficult to link them together. To go further, we must therefore break the rigid constraints of these administrative data. At the same time, however, when the information becomes too detailed, we can no longer use a comprehensive database but we must content ourselves with observing a sample. Event-history surveys meet this dual requirement. We shall describe them here briefly, referring the interested reader to the multi-author volume by the "Groupe de Réflexion sur l'Approche Biographique" (1999).

Event-history surveys must register a maximum number of events occurring in the lives of the persons interviewed, by chronological order of occurrence and noting the intervals between them, so as to identify all their potential interdependencies. They must also record the largest possible number of characteristics of respondents and their living conditions in order to incorporate them into the analysis of behaviors. To capture all these events and statuses, we can use two types of questionnaire. Each has its advantages and drawbacks.

The first type is the prospective survey that follows the members of a sample throughout their lives. Depending on the survey's more specific objective, it can start at the person's birth or at any other life stage that is useful for the analysis—for example, at marriage to study legitimate fertility, union terminations, remarriage, and so on. We can then visit the sample annually, for example, to record a small number of events or changes in characteristics. In this case, the event-dating accuracy will be excellent. On the negative side, the time between the start of the survey and the data analysis, which cannot begin until after a sufficiently long period, tends to discourage many researchers. Also, the expected results are biased by the loss of individuals during the survey ("attrition"), essentially due to migration to a location that the interviewer cannot determine, or to the refusal to continue to respond. The best example of such a survey is the one carried out by Cribier and Kych (1999) on a sample of persons who were retired at the time the survey began.

Alternatively, the researcher will often prefer to conduct a retrospective survey, recording in a single session all past events of interest and individual characteristics. The survey is thus immediately usable for analysis, with no attrition risk. By contrast, such a survey is a far heavier operation (Courgeau, 1999a). We should, however, note the risk of bias due to the survival-based selection of respondents, which is generally minimal (Lyberg, 1983). Respondents' dating errors and failure to recall events may also be significant. We have had the opportunity to test these memory errors in Belgium, a country that maintains population registers. We were able to show that the errors concern the exact dating of events (Poulain et al., 1991), but do not alter the logical order of events, which is recalled accurately (Courgeau, 1991b). The survey is therefore reliable in those areas where the analysis requires it to be.

Whatever the type of survey, some of the inter-event intervals will not be fully observed. The observations are interrupted ("censored") by the survey date in retrospective surveys, or by the end of the observation in prospective surveys, unless the survey continues until all respondents are extinct. Actually, we know how to use this information, which tells us that the individual did not experience the event studied before the observation.

The first analyses concerned the observation of a single phenomenon, which authors sought to explain by various individual characteristics: Menken and Trussell (1981) analyzed

the dissolution of marriage as a function of selected socio-demographic characteristics of respondents. The analyses were then applied to more complex interactions between different demographic events, while taking into account the heterogeneity of the population under study.

Our purpose here is not to go into the details of the analytical methods, but simply to note the modus operandi; we refer the interested reader to works of a more mathematical nature (Andersen et al., 1993) or more demographic nature (Courgeau and Lelièvre, 1989, 1992, 2001). In the following pages, we shall discuss the case of two interacting demographic phenomena—marriage and migration, fertility and migration, etc.—that are also determined by various individual characteristics, time-dependent or not. We begin by analyzing the interaction between the phenomena without bringing the individual characteristics into play.

The principle of this method is to define two random variables T_1 and T_2, corresponding to the time elapsed between the initial observation instant and the occurrence of the events. We can then estimate instantaneous hazard functions of occurrence of each event depending on whether the other has already occurred or not (for more details, see Courgeau and Lelièvre, 1989, 1992, 2001). We can thus write:

$$h_{01}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(T_1 < t + \Delta t | T_1 \geq t, T_2 \geq t) \qquad \text{(III.10)}$$

which, when Δt tends toward zero, gives us the instantaneous hazard of the first event when the second has not yet occurred. Likewise:

$$h_{21}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(T_1 < t + \Delta t | T_1 \geq t, T_2 = u) \quad \text{where} \quad u<t \qquad \text{(III.11)}$$

which, when Δt tends toward zero, gives us the instantaneous hazard of the first event when the second has already occurred. This hazard should also depend on the date of the second event's occurrence, u; for simplicity's sake, however, we assume here that the dependence is negligible.

It is easy to write the two hazard functions for the second event in symmetrical form. We know how to estimate these functions when some observations are interrupted, as well as their variance, which enables us to test for the equality of some of the functions. The tests can also reveal interesting dependencies between the two events. For instance, one of the phenomena may depend significantly on the previous occurrence of the other, which, conversely, does not depend on the first: this is called unilateral or local dependence (Schweder, 1970) of the first on the second. This concept of unilateral dependence is far closer to the concept of causality—even though we can never strictly speak of causality in social science—than the correlations measured in period analysis. By introducing time into the picture, it signals that the occurrence of a phenomenon may alter the hazard of experiencing the other. Conversely, a correlation does not signify causality, as the two phenomena may depend on a third, unmeasured phenomenon while being independent of each other.

Naturally, we may observe reciprocal dependence between two phenomena, when each significantly influences the other. Total independence between the events is also possible, but this case seems far rarer. Now this finding challenges precisely one of the conditions of applicability of cohort-analysis methods, namely, independence between events.

Other, more complex dependencies may be identified. For example, in studying interactions between fertility and migration to metropolises (Courgeau, 1987), we observe a sizable reduction in migrant women's fertility of order higher than 1. The question is the following: is this an instance of adaptive behavior (in which migrant women copy the low fertility of urban women) or selective behavior (in which future migrant women already behave differently from other women in the departure area)? We observe that the fertility of future female migrants to metropolises is already lower than that of sedentary women in weakly urbanized areas, and that this fertility is identical to that of women who have already migrated to metropolises. We have thus identified an a priori dependence of fertility on future migration, expressed by this selection in the initial population.

Equally, non-parametric analysis can reveal dependencies that change with age. For instance, young women born between 1911 and 1936 who have had a first child and returned to work are more likely to have a second child than those who have not gone back to work. However, after thirty years, we observe the opposite: these older women are less likely to have a second child than those who have not started working again (Lelièvre, 1987).

As noted previously, while the Norwegian data allow a complete tracking of migrations by the population observed, the source unfortunately does not record exits from agriculture. It is therefore impossible to study the links between the two phenomena in Norway. INED's retrospective survey on "Family, Occupational, and Migration Biography" (also known as "Triple Biography" or 3B) offers a life-long tracking of nuptiality and exit from agriculture for French persons born between 1911 and 1936. We shall use it in the following example.

Example no. 4: Nuptiality and agriculture in France

This example will be briefly described here; for further details, we refer the interested reader to the article by Courgeau and Lelièvre (1986). Let us focus on the 519 women in the sample who began their working life in agriculture. We calculate (1) the cumulative hazards of nuptiality at each age starting at 15 years, according to whether the women stayed in agriculture or exited; (2) the hazards of exit from agriculture for the same women according to whether they are still unmarried or already married, at the same ages. The two series of hazard functions are plotted in charts III.3 and III.4.

The examination of both charts and the tests, which show whether the differences between the curves of each of the two series are significant, yield very clear results. For women, we detect no influence of exit from agriculture on nuptiality, at any age. The slight differences that seem to appear between ages 35 and 40 are absolutely not significant. By contrast, once married in the agricultural world, they will remain in it far longer than their unmarried colleagues, with an unquestionably significant difference. This unilateral dependence shows us a strategy where marriage can allow women to stay in agriculture. Marrying a farmer may arguably give them access to a larger farming concern. Our semi-parametric analysis below will shed light on these behaviors.

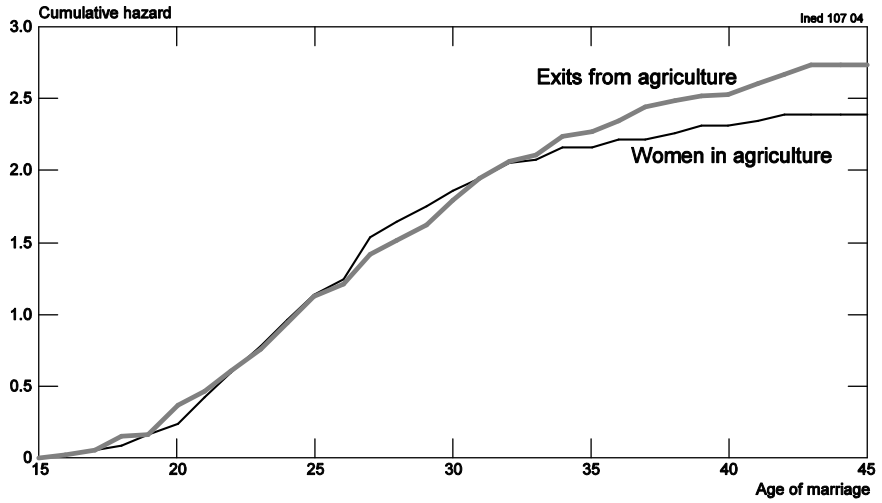


Figure III.3. - Cumulative hazard functions of nuptiality for women remaining in agriculture and those who have exited

Figure III.4 – Cumulative hazard functions of exit from agriculture for unmarried and married women

A symmetrical analysis of the male population (Courgeau and Lelièvre, 1986) also shows a unilateral dependence, but in the opposite direction: marriage has no impact on male farmers' exit from agriculture, but their exit from agriculture boosts their low probability of marrying by two-thirds.

Just as we have introduced dependence between phenomena in an event-history model, so can we introduce population heterogeneity in a generalization of these non-parametric models. To perform this function, the logistic models used earlier require the introduction of time. This has been possible in parametric and especially semi-parametric event-history models, described in greater detail below. The proportional-hazard model developed by Cox (1972) is the most commonly used in demography, although other types exist such as the accelerated

failure time (AFT) model. With the Cox model, we can write the instantaneous hazard, already defined in the non-parametric case, as:

$$h(t;z) = h_0(t)\exp(\beta z) \qquad \text{(III.12)}$$

where z is a characteristics vector with a multiplying effect on the underlying instantaneous hazard $h_0(t)$.[9] It is easy to see that when a characteristic is binary, the hazard of an individual who possesses it is equal to that of an individual who does not multiplied by a constant term $exp(\beta)$, hence the designation "proportional-hazard model."

We know how to estimate the parameters of these models using partial likelihood, as well as the underlying instantaneous probability (see Courgeau and Lelièvre, 1989, 1992, 2001). Naturally, we shall need to verify beforehand that such a model does indeed apply to the phenomenon studied: for example, for binary characteristics, we conduct non-parametric analyses on each to confirm the suitability of a multiplicative model. We plot the curves for the logarithms of the cumulative hazard functions of individuals with or without the characteristic studied, as a function of time: the curves should be approximately parallel.

Next, we simultaneously introduce population heterogeneity and interaction between the demographic phenomena. We can do this with generalized Cox models, as they can incorporate time-dependent characteristics that enable us to measure the occurrence or non-occurrence of interacting phenomena. Consequently, if we want to pursue the analysis of the interaction between two phenomena begun earlier, we can write in a more summary form the instantaneous hazard of occurrence of one of the events, according to whether or not the other occurred previously at the instant u, as follows:

$$h_{.1}(t;z,z',u) = h_{.1}^{0}(t)\exp[z\beta_1 + H(t-u)(\beta_0 + z'\beta_2)] \qquad \text{(III.13)}$$

where:

$$H(x) = \begin{cases} 0 & if \quad x < 0 \\ 1 & if \quad x \geq 0 \end{cases}$$

u is the date of occurrence of the other event, z and z′ the characteristics influencing the phenomenon studied before and after the occurrence of the other event. Some of these characteristics may remain the same over time; others may appear or, on the contrary, cease to have an influence after the occurrence of the other event. Here as well, we can still estimate the parameters by means of maximum partial likelihood.

Let us return to the previous example to see the effect of various characteristics, both fixed and dependent on the other event, on exit from agriculture by Frenchwomen born between 1911 and 1936.Example no. 4 (continued): Nuptiality and agriculture in France

We continue to view a woman's exit from agriculture as dependent on her marital status but also on a certain number of characteristics that remain constant over time (number of siblings, eldest child, and farmer father), and other characteristics acquired at marriage (farmer husband, farmer father-in-law, and woman still in agriculture at the time of her marriage). The estimated parameters are reported in table III.6.

[9] We can also define parametric models where $h_0(t)$ is a given function of time, determined by a certain number of parameters to be estimated. Various distributions have been proposed: exponential, Gompertz, Weibull, log-logistic, etc. (for more details, see Courgeau and Lelièvre, 1989).

TABLE III.6. - ESTIMATED PARAMETERS OF SEMI-PARAMETRIC MODEL OF WOMEN'S EXIT FROM AGRICULTURE

Set of characteristics	Main effect β_1	Interference β_0	Interaction β_2
Number of siblings	0.012**		0.000
Eldest child	-0.320**		0.296
Farmer father	-0.928**		0.806*
Married		-0.228	
Farmer husband			-0.359**
Farmer father-in-law			-0.126
Farmer at marriage			-1.040

* Significant at 10% level.

** Significant at 15%level.

We can see that the more siblings a woman has, the more likely she will be to exit from agriculture. Conversely, women who are the eldest children and women with a farmer father are less likely to exit from agriculture. But once married, while the effect of the number of siblings does not change, the fact of having a farmer father no longer influences the probability of exiting from agriculture. We also see that having a farmer husband will keep the woman in agriculture, corroborating our hypothesis on the reasons why women who marry in the agricultural community stay in it.

Because the model is multiplicative, we can compute the risk of exit from agriculture for women with given characteristics. For example, a woman who is an only child (no siblings) and whose father is a farmer has a risk of exiting from agriculture before marriage of exp(-0.320 - 0.928) = 0.287, whereas a woman with 10 siblings and a non-farmer father will have a risk of exp(0.12) = 1.127, or nearly four times as high as that of the first.

These various applications of event-history methods yield conclusions that seem more robust than those obtained with period analysis or cohort analysis. We can thus formulate more clearly the paradigm underlying the event-history approach to display its advantages with respect to the previous methods, but, at the same time, to pinpoint the aspects that it too leaves in the shadows.

II. - Paradigm of event-history approach

The focus of our study has now shifted away from homogeneous sub-populations toward a set of individual trajectories between a large number of states. The analytical unit will no longer be the event but the individual event history, viewed as a complex stochastic process. We no longer seek to isolate each phenomenon in its pure state, but, on the contrary, to see how an event in a person's life can influence his or her later life course, and how certain characteristics can lead one person to behave differently from another.

The paradigm, in this case, can be broadly defined as follows: individuals follow complex, life-long trajectories that depend, at a given instant, on their earlier trajectories and on the information that they have acquired in the past (Courgeau and Lelièvre, 1996). In other words, this is a resolutely individualistic approach that reflects a methodological individualism (Valade, 2001) and shows that people's behavior is connected to their prior life histories, without seeking the motives for their acts in society. It is therefore diametrically opposed to the aggregated period approach, which, as we saw, is a methodological holism.

To begin with, we shall follow over time a set of individuals belonging to a given generation or cohort. The main way for an individual to escape observation is to exit from the sample at the survey date or study date, if we are working in a prospective framework. Insofar as there is no reason for these dates to be linked to an individual's life, the independence condition is totally fulfilled: the observation is called "non-informative" and a method exists to take account of these exits when estimating probabilities. We can thus see how a change in viewpoint solves the problem of disturbing phenomena.

On the other hand, there is a risk of selection bias, particularly in retrospective surveys, due to the fact that we can only interview survivors present in the country at the time of the survey. We are very often forced to assume that the exit from the population studied is non-selective, unless register data are available to correct these biases (Hoem, 1985). The biases are, however, lessened if the event studied does not occur in an elderly population or a population with a heavy emigration rate.

We can work on sub-populations—living in or outside metropolises, for example—and study the occurrence of events such as successive births (Courgeau, 1987). If these persons experience disturbing phenomena (exit from initial residence environment), they no longer exit from the study's scope of coverage. On the contrary, they can adopt a different fertility behavior. We can test this change of behavior by comparing it with that of sedentary persons of the same age, or that of the sub-population to which they have migrated. This enables us to determine, as noted earlier, whether we are dealing with a possible selection or, instead, with behavioral adjustment.

Whereas in classic cohort analysis there was no reason to distinguish between disturbing events and competing events, we must now examine them separately. As pointed out above, the disturbing phenomenon—which we prefer to call "interacting phenomenon" here—alters the probability of occurrence of the event studied. By contrast, when we speak of competing phenomena, we refer to different variants of an event that produce the same outcome: mortality classified by cause; union formation classified according to whether it is due to marriage or cohabitation, and so on.

But the goal here is not to try to determine mortality for a given cause, on the assumption that the other causes are eliminated, nor to calculate the hazard functions of nuptiality in the absence of cohabitation. These questions lie outside our statistical field, and

the answers that social science can attempt to offer should be viewed with the greatest circumspection. We shall show, instead, how these different causes act simultaneously, how the exit from never-married status takes place via marriage or cohabitation, without trying to separate these effects. In other words, we no longer define an intensity of the exit toward these different statuses, but only compare the hazard and timing of each exit.

We are thus equipped to see how a family event, economic event or other type of event experienced by an individual with a known past experience will alter the hazard functions of occurrence of the other events in his or her life. We shall seek to determine, for example, how marriage can influence a person's working career, spatial mobility, the occurrence of other events, such as the birth of a child, a break with the family of origin, and so on. This effectively constitutes what we have called the analysis of interactions between demographic phenomena, which fits comfortably into the proposed paradigm.

Such an analysis presupposes an initial population that is homogeneous with regard to the process studied, i.e., at the start of the analysis, all the individual trajectories are at the same stage of the process. But the population becomes heterogeneous over time, as it experiences the various interacting events. This hypothesis can be used at the beginning of the analysis to untangle the interactions between the phenomena, but it must be waived later on. Indeed, there is no reason why the members of the initial population should be identical. Time regression methods, used in the second phase of the analysis, allow an exploration of the population's initial heterogeneity and the heterogeneity that develops in the following periods.

If we want to understand a person's behaviors, we will need to refer to his or her social origins and entire life history. We accordingly assume that these behaviors are not innate, but change during an individual life course thanks to personal experiences and successive acquired factors. For instance, two persons born in families that are initially very close because of their common social, religious, and occupational origins, etc., but having followed divergent careers, will display behavior toward different demographic phenomena that will also diverge over time.

We then arrive at the analysis of population heterogeneity, now seen in dynamic form and not in static form as in period analysis. The regression analysis used in the period approach sought to connect aggregated behaviors with characteristics of equally aggregated populations. We must now extend it to the analysis of individual characteristics. When a person is born, his or her life can follow a wide variety of paths. Despite this, the paths are far from being all equally probable. Accordingly, an individual's life history is defined as the result of a complex stochastic process, unfolding in his or her life course.

These processes have been studied by probabilists and statisticians with the aid of martingale theory (Dellacherie and Meyer, 1980), stochastic integration (Dellacherie, 1980), and counting processes (Brémaud and Jacod, 1977). Space precludes a description of these methods here, and we refer the interested reader to the excellent overview by Andersen et al. (1993). We shall simply indicate the outlines of the approach to show how it optimizes the use of the proposed paradigm.

It describes an individual life as a stochastic process, unfolding in a given generation or cohort. These individuals may experience a number of demographic events that cause them to shift from one state to another. Naturally the order of occurrence of the events is not specified and can be very variable. But it is equally obvious that the chance of experiencing one of these events at a given moment will be linked to the person's earlier history (known events, their order, and the dates of their occurrence) and characteristics that are fixed (social and

geographic origins, number of siblings, and so on) or variable over time (economic crises, wars, etc.).

To study these transitions, we shall use a basically semi-parametric model, which will model parametrically the effect of the characteristics of interest in our research without making any assumption about the type of distribution of the length of stay in each state studied. The model will be dynamic, as it enables us to model the instantaneous rates of occurrence of the events studied for the various populations at risk. It also offers the prospect of introducing time-dependent characteristics and thus allows the estimation of a truly dynamic model of the evolution of stochastic processes over time. These changes will be linked essentially to individual events and characteristics.

Another important property of these models is their ability to incorporate interaction effects between individual characteristics. For example, if migration, at a given moment, depends on whether the person is a farmer and on his or her marital status, there can be a difference in behavior between never-married farmers and the others, as well as between married farmers and the others. We can add to this interaction effect by introducing the product of the binary variables corresponding to the two characteristics and we can estimate the model that incorporates the three variables. Naturally, we can generalize this logistic model into an event-history model, which can include the same interaction between characteristics, now time-dependent.

However, the model displays little sensitivity to several issues such as exit from observation at the time of the survey (a factor that we know how to take fully into account), the occurrence of competing events, or the existence of unobserved heterogeneity. Let us examine the last point in greater detail, as it is crucial to the validity of the analysis performed.

When we conduct an analysis, we can be sure that it will be unable to include all the factors influencing the process studied: some are not captured by the survey, while others are regarded as having no effect whereas they actually do. This creates what is known as unobserved heterogeneity, which can invalidate the results obtained with the observed data alone. We know that when we analyze period data with regression models, this unobserved heterogeneity leaves the estimated parameters unchanged if we include it when it is independent of the observed variables. This is unfortunately not the case when we use a semi-parametric model that incorporates time.

However, Bretagnolle and Huber-Carol (1988) were able to study the effect of unobserved characteristics on the parameters estimated with the observed characteristics. They showed that, when the omitted characteristic is independent of the observed characteristics, the omission has no effect on the sign of the estimated parameters. On the other hand, it introduces a reduction in the parameters' absolute values. This means that, if the effect of a characteristic seems fully significant when the other characteristic is omitted, the fact of introducing it into the model will only strengthen the effect of the first characteristic. In contrast, a characteristic with no significant effect on the phenomenon studied may acquire a fully significant effect when we introduce the unobserved characteristic.

These results are very important as they enable us to be certain of the observed effects' direction and significance, even though we do not know if we have introduced all the characteristics influencing lengths of stay into the model. When presenting cohort analysis, we showed how heavily this problem weighed on the validity of such an approach. We now have a means to verify its effect on the results of an event-history analysis.

Although the event-history approach offers many advantages over the period approach as well as the cohort approach, it still leaves unresolved a number of issues that we shall now examine.

III. - Problems encountered with this approach

While this approach provides more effective responses to some of the issues raised in the previous chapters, it does pose problems of its own with regard to specific points relating to event-history analysis of human data. In the section above, we addressed issues regarding unobserved heterogeneity. We must now discuss in greater detail some other problems that it creates in event-history analysis, before examining the risk of atomistic fallacy that such an analysis entails.

Unobserved heterogeneity

We have already shown in detail how to introduce population heterogeneity into an event-history analysis and the effect of characteristics that are unobserved—but independent of observed characteristics—on parameters estimated with a Cox model. It will be interesting, however, to examine some other possibilities more fully.

Let us begin with the case where we initially observe only a single characteristic, which seems to have an unquestionably significant effect but can lose it entirely once we simultaneously observe other characteristics that are more relevant yet not independent of the first. In this case, we can say that the first characteristic was a variable that was simultaneously subjected, but in a less precise manner, to the effect of the characteristics introduced later on. Once these are introduced into the model, the variable no longer has any significant effect. To see more simply how this can happen, let us describe a clear example of such effects.

Example no. 5: Migrations of French males born between 1931 and 1936

The "3B" survey enables us to analyze pre-1981 migrations (changes in dwellings) by several cohorts of French males (for more details, see Courgeau, 1985). Let us concentrate here on migrations by men born in the period 1931-1936. We model the length of residence in each dwelling by incorporating a growing number of individual characteristics provided by the survey.

All migration studies have consistently shown a strong effect of age on the migration rates observed (Willekens and Rogers, 1978; Courgeau 2003c) and we can easily verify that this pattern applies to the generation observed (solid curve in figure III.5), with a peak at 20-24 years. Arguably, however, we can explain this age effect more precisely by the entry of cohort members into different life stages, which occur at equally different ages. For instance, once we include different stages of family life (marriage, divorce, widowhood, birth of children, departure of first child, etc.), we see that the age effect diminishes between 15 and 25 years of age, for it is taken into account by these events (dotted curve in figure III.5). If we now include housing tenure status (housed by parents, tenant, owner-occupier, or housed by employer), the age effect is further reduced, and sharply, for all ages between 20 and 35 (broken curve in figure III.5). But when all characteristics—concerning family, tenure status, occupation, and political or economic circumstances (war, economic crisis, etc.)—are taken into account, the age effect vanishes entirely, as the last curve shows. Age ceases to have any significant effect, despite the fact that the model is of far better quality than the first: all the other events fully explain the initial age curve.

This analysis effectively shows that the population heterogeneity that seems due to age is fully explained by successive life stages, which substantially modify the hazard functions of migrating.

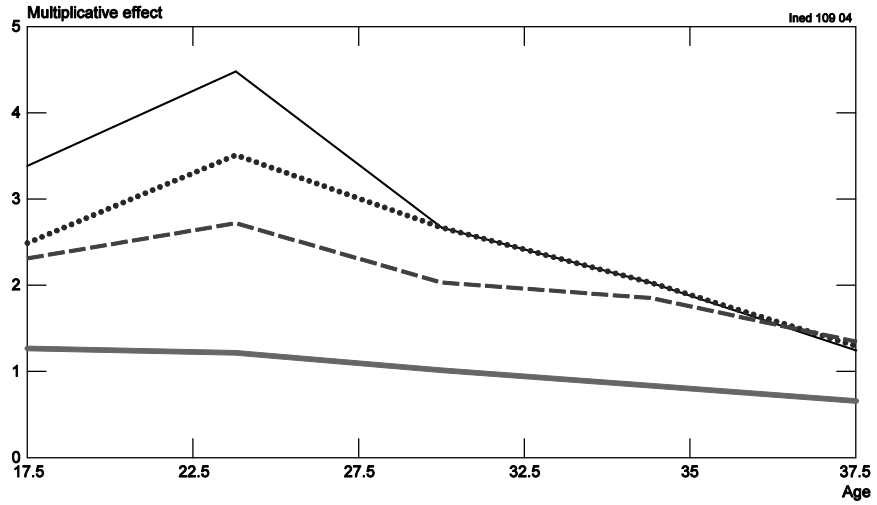


Figure III.5. - Multiplicative effect of age on mobility of cohort born between 1931 and 1935, as a function of selected characteristics taken into account

Alternatively, after introducing all the observed characteristics, some authors have tried to model unobserved heterogeneity (Vaupel et al., 1979; Manton et al., 1992). As noted in III.1, the hypothesis of an identical probability for all individuals whose characteristics have a given value—a hypothesis we need to make to estimate parameter variance—is rather unlikely. It can be tempting to introduce this heterogeneity in the form of a distribution of a given type, also called frailty, for it can interfere with the parameters of the characteristics observed. Earlier, we saw that when we use a Cox model, this effect is equivalent to underestimating the parameters' absolute values. The effects can be even more disastrous with a parametric model: at worst, some parameters change signs (Trussell and Richard, 1987). However, for the analysis of repetitive events, we can show that while there is only one model without unobserved heterogeneity, there exists an infinity of models with unobserved heterogeneity and estimated hazard functions that differ but adjust identically to the observed data (Trussell and Rodriguez, 1990). We can choose between these models only on the basis of information about heterogeneity, which is generally unavailable. We therefore believe that, in this case, the choice of an arbitrary distribution for heterogeneity does not solve any problem. Only new information of a more biological nature—in contrast to the social or cultural information already available—would allow a choice between the proposed models (for more details, see the introduction by Trussell [pp. 4-6] to the work by Trussell et al., 1992).

By contrast, if we analyze repetitive events, we have an opportunity to estimate fuller multilevel models that can introduce unobserved heterogeneity, reflecting the multiple events observed for an individual (Lillard, 1993). We shall explore this possibility in more detail in the next chapter.

Risk of atomistic fallacy

We can now identify the factors at work—both demographic and non-demographic—and analyze their effects on individual behaviors in great detail. As a rule, however, it is the characteristics of individuals themselves that will be invoked to explain their behaviors. In

doing so, we may commit what is known as the atomistic fallacy, for we overlook the context in which human behaviors occur. This context may be defined as the family environment in which the person lives, or, more generally, as a relatively wide contact circle around the person: neighborhood, town, and so on. Arguably, this context can influence individual behaviors, and it seems fallacious to isolate individuals from the constraints of the society and environment in which they live.

This fallacy risk, which we can contrast with the ecological fallacy, had already been recognized by sociologists (Lazarsfeld and Menzel, 1961). They showed the need to define with sufficient precision different types of groups, communities, organizations, and, more generally, any sets of individuals. These sets may be composed of members who are comparable in terms of the behavior studied and who must be described by a certain number of properties. An entity treated as a group in one study may be viewed as a member of a broader grouping in another study. This property is very important, for it shows the relativity of the individual, which the event-history approach regards as the dominant unit. Cautiously applied, the approach should enable us to transcend the opposition between individualism and holism.

We shall now see how that can be done.

CHAPTER IV

TOWARD A CONTEXTUAL AND MULTILEVEL ANALYSIS

The risk of atomistic fallacy incurred when working with event-history data contrasts with the risk of ecological fallacy associated with aggregate data. Avoiding both risks would require the ability to work at different levels simultaneously.

This is a difficult undertaking, as the event-history approach predicts an individual behavior with the aid of equally individual characteristics, while the aggregate approach predicts a collective behavior with the aid of the group's characteristics. One way to escape this dilemma is to try to predict an individual behavior with the aid of both individual characteristics and group characteristics.

If we choose this option, we can include different measures of the same characteristic, i.e., each individual can be associated both with the fact of possessing a given characteristic (individual measure) and with the fact that the group to which (s)he belongs is characterized by the mean characteristic of its members (aggregate measure). Thanks to this contextual approach, we can not only identify the sources and size of the ecological bias, but also clearly separate the individual, contextual, and ecological effects that influence the phenomenon studied.

However, we need to go even further and implement a true multilevel analysis, which introduces an internal dependence in each group (ignored by contextual analysis), while continuing to take account of individual and contextual characteristics simultaneously. This analysis will require new methods to estimate the effects of the characteristics and the random effects operating at each aggregation level. The present chapter describes the essentials of this new approach, which will be treated in greater detail in Part II.

I. - ESTABLISHING THE ANALYSIS

The multilevel approach was introduced into demography shortly after event-history analysis, toward the mid-1980s (Mason et al., 1983). We begin by describing contextual

analysis—which simply generalizes classic methods—before addressing multilevel analysis in all its complexity.

From contextual analysis...

First, we can see more precisely how the ecological fallacy and the atomistic fallacy can affect the analysis of a specific type of data, when we use the data in aggregated form or, on the contrary, in individual form. Let us take the example of an analysis of binary data in the form of a logistic model, while bearing in mind that we can perform an identical analysis on continuous data (Firebaugh, 1978) as well as on event-history data (Baccaïni and Courgeau, 1996).

Let us rewrite model (III.5) of chapter III, examining only one explanatory characteristic but specifying the fact that individuals may be located in a certain number of areas j, although initially this presence will not influence the results:

$$P(y_{ij} = 1 \mid x_{ij}) = (1 + \exp - [a_0(1 - x_{ij}) + a_1 x_{ij}])^{-1} \qquad \text{(IV.1)}$$

The probability of experiencing the event studied, for the individual i who has this characteristic and is present in area j, can be written:

$$P(y_{ij} = 1 \mid x_{ij} = 1) = (1 + \exp - a_1)^{-1} = p_1 \qquad \text{(IV.2)}$$

and the probability for the individual who lacks the characteristic:

$$P(y_{ij} = 1 \mid x_{ij} = 0) = (1 + \exp - a_0)^{-1} = p_0 \qquad \text{(IV.3)}$$

As we can see, these probabilities, for the moment, are independent of the individuals' areas of residence. From these probabilities, we can determine the mathematical mean of the total number of individuals having experienced the event in area j as the sum of the probabilities of experiencing the event for each individual in the area:

$$n_j \bar{y}_{.j} = \sum_i P(y_{ij} = 1 \mid x_{ij}) = p_0 n_{0j} + p_1 n_{1j} \qquad \text{(IV.4)}$$

where $\bar{y}_{.j}$ is the estimation of the probability of experiencing the event in area j under logit-model conditions, n_{0j} is the population of individuals lacking the characteristic in area j, and n_{1j} the population of individuals possessing it. We can see that equation (IV.4) may be rewritten as:

$$\bar{y}_{.j} = p_0(1 - \bar{x}_{.j}) + p_1 \bar{x}_{.j} = p_0 + (p_1 - p_0)\bar{x}_{.j} \qquad \text{(IV.5)}$$

for the relationship $\frac{n_{1j}}{n_j} = \bar{x}_{.j}$ is indeed the proportion of individuals who have the characteristic and live in area j. The relationship we have just set up between the aggregate characteristics $\bar{y}_{.j}$ and $\bar{x}_{.j}$, whose parameters are estimated on individual data, is not necessarily the best possible fit obtainable from aggregate data. A population-weighted regression, of the type estimated in chapter I, will yield a better linear fit, in least-squares terms, and leads to the following relationship:

$$\bar{y}_{.j} = \gamma_0 + \gamma_1 \bar{x}_{.j} \qquad \text{(IV.6)}$$

The condition in which both types of model will lead to identical results is the equality of the parameters $(p_1 - p_0)$ and γ_1, the first estimated with a logistic regression on individual data, the second with a linear regression on aggregate data.

For a closer examination, we must now introduce both individual and aggregate characteristics into the same logit model. We can then write the corresponding contextual model as:

$$P(y_{ij} = 1 \mid x_{ij}, \bar{x}_{.j}) = [1 + \exp - (a'_0(1 - x_{ij}) + a'_1 x_{ij} + a'_2 \bar{x}_{.j} + a'_3 x_{ij} \bar{x}_{.j})]^{-1} \quad \text{(IV.7)}$$

where the parameter a'_2 stands for the aggregate characteristic and the parameter a'_3 represents the term for the interaction between individual and aggregate characteristics. As we can easily see, this probability now depends on the area in which individuals reside. This leads to the following probability for individuals possessing the characteristic:

$$P(y_{ij} = 1 \mid x_{ij} = 1, \bar{x}_{.j}) = [1 + \exp - (a'_1 + \{a'_2 + a'_3\}\bar{x}_{.j})]^{-1} \quad \text{(IV.8)}$$

and for those who do not:

$$P(y_{ij} = 1 \mid x_{ij} = 0, \bar{x}_{.j}) = [1 + \exp - (a'_0 + a'_2 \bar{x}_{.j})]^{-1} \quad \text{(IV.9)}$$

In consequence, the percentage of individuals having experienced the event in area j will no longer be a linear function of the proportion of individuals who have the characteristic, for we can, as earlier, write:

$$y_{.j} = (1 - \bar{x}_{.j})[1 + \exp - (a'_0 + a'_2 \bar{x}_{.j})]^{-1} + \bar{x}_{.j}[1 + \exp - (a'_1 + \{a'_2 + a'_3\}\bar{x}_{.j})]^{-1} \quad \text{(IV.10)}$$

As we can see, for the relationship (IV.10) to be equivalent to the formula (IV.5), it is necessary and sufficient for a'_2 and a'_3 to be null. In other words, the aggregate characteristic must have no effect on the probability of experiencing the event, when we are simultaneously checking for the effect of the individual characteristic. In this case, the aggregate and individual models are perfectly equivalent. The problem stems from the fact that this situation is rarely observed in reality, as already noted in the Norwegian example.

In the general case, we therefore need to bring data measured at the individual and aggregate levels into play simultaneously to explain a behavior that remains individual. The simplest solution to this problem is to use data measured at different aggregation levels to explain an individual behavior. We can now grasp the difference between this approach, which uses aggregate characteristics to explain an individual behavior, and the aggregate approach, which explained an aggregate behavior by equally aggregate characteristics.

We can thus eliminate the risk of ecological fallacy, for the aggregate characteristic will measure a different construct from its equivalent at the individual level. It no longer acts as a substitute, but as a characteristic of the sub-population that will influence the behavior of a member of that sub-population. Simultaneously, we remove the atomistic fallacy, as we take into consideration the context in which the individual lives. We may well ask, however, if the inclusion of the aggregate characteristics provides an entirely sufficient representation of that context: as we shall see later, it will be necessary to take further steps in a fully multilevel analysis.

For the moment, let us describe a specimen application of this type of contextual model.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

Let us go back to the example of Norwegian farmers by incorporating simultaneously into the model the proportions of farmers residing in each region. The estimated values of the parameters of model (IV.6) are listed in table IV.1.

TABLE IV.1 - ESTIMATION OF PARAMETERS OF CONTEXTUAL LOGIT MODEL WITH SIMULTANEOUS INCLUSION OF INDIVIDUAL AND AGGREGATE CHARACTERISTICS

Characteristics	Parameters of logit model	
	Estimated value	Standard deviation
Non-farmer	-1.996	0.033
Farmer	-2.155	0.209
Proportion of farmers	4.469	0.461
(Farmer x proportion of farmers) interaction	-5.774	2.447

The first clear lesson is that the effects of the aggregate characteristics are entirely significant—hence the need for a contextual model here. With the model's results, we can harmonize the contradictory results obtained previously: the fact of being a farmer still sharply reduces a person's probability of migrating, while the fact of inhabiting a region with a high percentage of farmers will increase the likelihood of migrating for non-farmers and farmers. But we find a totally different explanation for why emigration rates rise in tandem with the percentage of farmers. In the aggregate-data analysis, it is the constancy of the probabilities of emigrating for farmers and non-farmers (with different values) that—via the differing percentages of farmers—explains why emigration rates are highest in regions with the most farmers. Here, the same result is explained by the variation in those probabilities of emigrating as a function of the percentage of farmers. Figure IV.1 shows that the higher probability of migrating is largely due to non-farmers, for farmers' likelihood of migrating is not significantly affected by the percentage of farmers. This model does support our earlier hypothesis that a high density of farmers in an area will increase the probabilities of migrating among other occupations. The possible explanation, in these regions, may be a relative lack of non-farm jobs, which drives persons in other occupations to emigrate more than farmers when they are seeking new work.

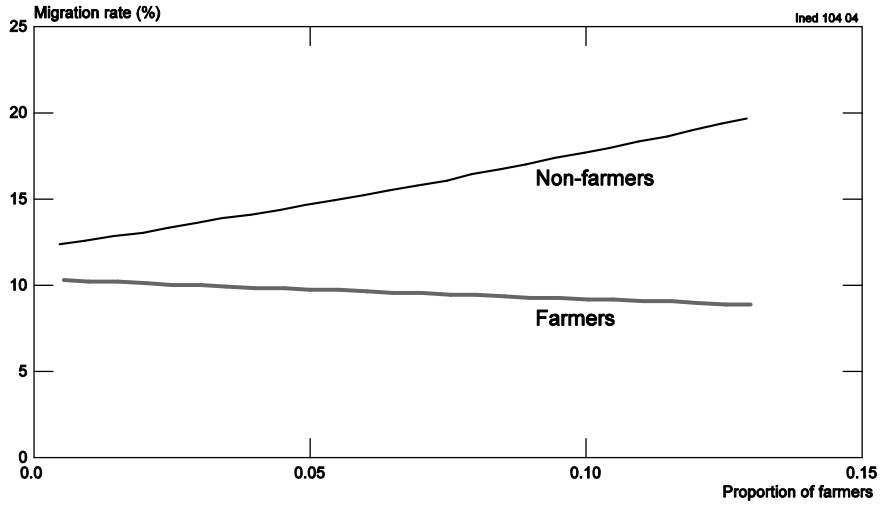


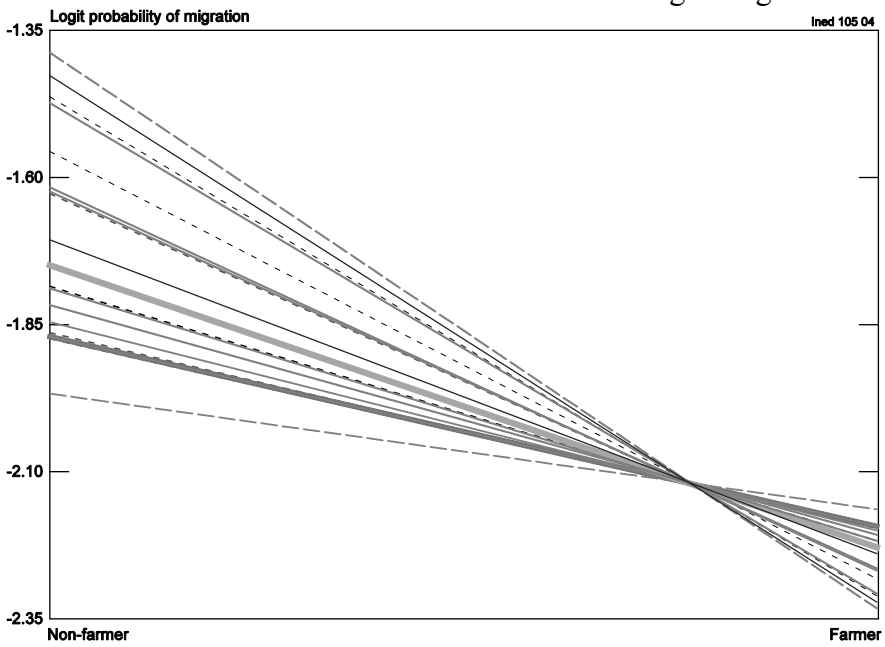
Figure IV.1. - Migration rate of farmers and non-farmers in Norway as a function of the proportion of farmers in each region

The use of contextual models imposes highly restrictive conditions on the formulation of the log-odds (logarithm of relative risks) as a function of characteristics. In particular the models assume that the behaviors of individuals within a group are independent of one another. In practice, the risk incurred by a member of a given group more likely depends on the risks encountered by the group's other members. Overlooking this intra-group dependence generally biases the estimates of the variances of contextual effects, generating excessively narrow confidence intervals. Likewise, these log-odds, for individuals in different groups, cannot vary freely but have restrictive constraints imposed by the model used (Loriaux, 1989). Let us see, for example, what happens in the previous example.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

If we link the log-odds of farmers' migration (x=1) to those of non-farmers' migration (x=0) in each region, we obtain a series of straight lines intersecting the points with coordinates (0, $a'_0 + a'_2 x_{.j}$) and (1, $a'_1 + (a'_2 + a'_3) x_{.j}$) respectively, as shown in figure IV.2. We can easily verify that all these lines pass through a common point with coordinates $(-\frac{a'_2}{a'_3}, a'_0 - (a'_1 - a'_0)\frac{a'_2}{a'_3})$. In fact, however, there is no reason for such a constraint to be fulfilled. As shown below, a multilevel model lifts the constraint altogether.

Figure IV.2. - Logits of probabilities of migrating for farmers and non-farmers, estimated with a contextual model for each Norwegian region



If we plot the risks of migrating instead of the log-odds of migrating on the y-axis, the convergence of straight lines, while no longer verified, remains fully satisfactory as shown in figure IV.3.

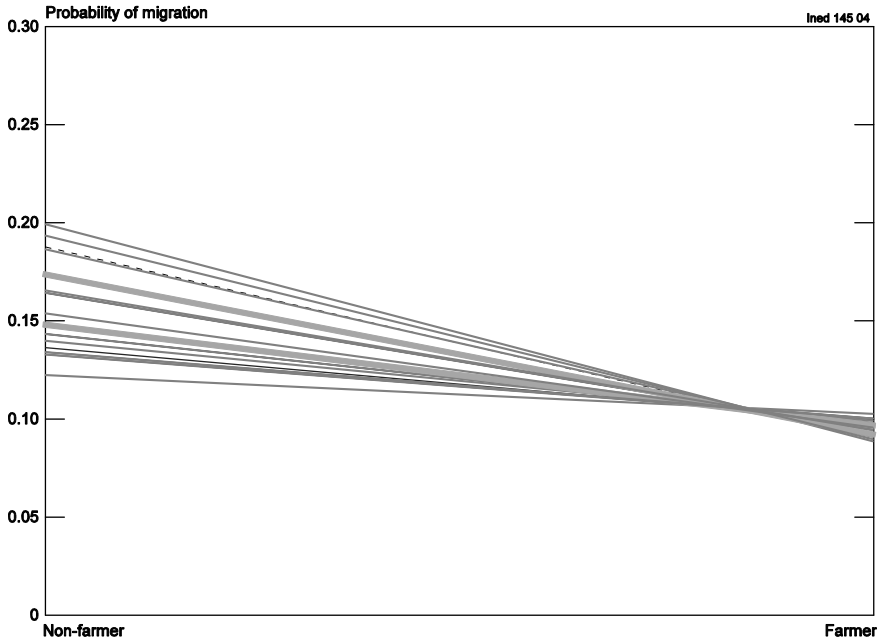


Figure IV.3. - Probabilities of migrating for farmers and non-farmers, estimated with a contextual model

These constraints make it necessary to formulate such a model in its most general form. We now arrive at true multilevel models.

…to a multilevel analysis

To illustrate our course of action more clearly, it is useful to revisit our earlier remarks on cohort analysis, when we were seeking to identify the effect of various characteristics on demographic behaviors. We noted that a decomposition into more homogeneous sub-populations for different regions as well as for different characteristics rapidly entailed very small numbers of persons at risk. The results of such an analysis became too unstable and could no longer reveal significant relationships. The analysis was drowned in random fluctuations (noise) that concealed any meaningful finding.

To solve this problem, we turned to linear-regression or rather logistic-regression demographic methods, which identified the salient results of the analysis. By introducing time into Cox models, we could study population heterogeneity and obtain results that were now totally significant, even when we included a large number of characteristics. But this method called for new hypotheses that needed testing. Unfortunately, the power of these tests is so weak that they virtually prevent us from rejecting the model even if it is severely flawed. Contextual models, which generalize these methods by incorporating different regions or groups of individuals, still offer no solution to the difficulties just noted.

At this point, it seems useful to seek a compromise between (1) a model that places no constraint on its estimators, but virtually precludes a significant estimation, and (2) a model with excessively strong constraints whose validity we can hardly test. The solution to this double problem lies, in our opinion, in multilevel models. These will introduce random effects into the previous individual or contextual models, so that we can generalize the regression methods discussed.

Let us go back to the model used throughout this chapter. Our first option is to estimate a logistic model for each zone j, in order to measure the effect of a characteristic on the probability of experiencing the event studied. We can accordingly write the model (IV.1) as follows for any individual i present in area j:

$$P(y_{ij} = 1 \mid i \in j, a_{ij}) = (1 + \exp - [\alpha_{0j}(1 - x_{ij}) + \alpha_{1j} x_{ij}])^{-1} \qquad \text{(IV.11)}$$

This is the first step in the analysis, which will estimate series of regional parameters. But when the number of regions or parameters is high, the parameters will be severely errored and will not enable us to draw meaningful conclusions. Hence the idea of constraining the parameters for more accurate results. If we assume, for example, that they are distributed at random around their mean value, we can write:

$$\alpha_{0j} = a_0 + u_{0j} \qquad \text{and} \qquad \alpha_{1j} = a_1 + u_{1j} \qquad \text{(IV.12)}$$

where u_{0j} and u_{1j} are zero-mean random variables. We can then restrict our analysis to the variances and covariances between these random variables:

$$\text{var}(u_{0j}) = \sigma^2_{u_0} \qquad \text{var}(u_{1j}) = \sigma^2_{u_1} \qquad \text{cov}(u_{0j}, u_{1,j}) = \sigma_{u_{01}} \qquad \text{(IV.13)}$$

Such a model therefore involves random variables at the individual level—as in any classic model—but also at the area level. It therefore requires special estimation methods presented in greater detail in the second part of this book. If we add the effect of aggregate characteristics to this model, it can be written in condensed form as:

$$P(y_{ij} = 1 \mid x_{ij}, x_{.j}) = (1 + \exp - [(a_0 + u_{0j})(1 - x_{ij}) + (a_1 + u_{1j}) x_{ij} + x_{.j}(a_2 + a_3 x_{ij})])^{-1} \qquad \text{(IV.14)}$$

Let us simply examine here how these various models apply to the Norwegian example.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

Let us first try to estimate a logit model for each Norwegian region to see in greater detail the regional variations in probabilities of migrating. Table IV.2 reports the estimates of logit-model parameters, the resulting estimates of farmers' and non-farmers' probabilities of migrating, and the test for the difference between these probabilities, for each Norwegian region.

The table shows that the differences between the probabilities of migrating are not significant at the 5% level except for 7 of the 19 regions. In the Oslo region, the difference between farmers' and non-farmers' probabilities of migrating is very large and carries the opposite sign to that of most other regions; the table shows that the difference does not, in fact, differ significantly from zero. As we can verify in table I.3 of chapter I, the region contains too few farmers for this result to be significant. Indeed, figure IV.4 shows very clearly the effect of the random factor on the estimated probabilities of migrating for farmers and non-farmers, which are extremely dispersed.

It is also interesting to compare this figure with figure IV.3, which gave estimates for these probabilities obtained with a contextual model. We can see clearly that the dispersion of regional results is much weaker in figure IV.3, especially for farmers: this confirms that the variances supplied by a logit model are drastically underestimated, as we feared earlier.

Hence the notion of using a fully multilevel model incorporating this inter-regional variance in addition to individual variance. The second column of table IV.3, entitled "Simple multilevel model," provides these estimates, which use equations IV.11 and IV.12.

We see that the new model's parameters differ little from those obtained with a simple logit model (see table III.3 of chapter III), but that their standard deviation is larger. This is because we no longer assume that all members of the population have the same probability of experiencing the migration; we now assume that the probability can change with the region. This is also shown by the random variables at the regional level, whose variance differs significantly from zero for non-farmers.

We can estimate the results of such a model for each region, which we can compare with figure IV.5.

TABLE IV.2.- PARAMETERS OF LOGIT MODELS, PROBABILITIES OF MIGRATING, AND TEST OF THE DIFFERENCE BETWEEN THESE PROBABILITIES (* IF DIFFERENCE IS SIGNIFICANT AT 5% CONFIDENCE LEVEL), FOR THE 19 NORWEGIAN REGIONS

Region	Logit-model parameter (standard deviation)		Probabilities of migrating (standard deviation)		Differences between these probabilities (standard deviation)
	Non-farmers	Farmers	Non-farmers	Farmers	
Ostfold	-2.148 (0.085)	-2.094 (0.975)	0.105 (0.008)	0.110 (0.095)	-0.005 (0.095)
Akerhus	-1.579 (0.059)	-1.145 (0.307)	0.171 (0.008)	0.241 (0.056)	-0.070 (0.057)
Oslo	-1.640 (0.047)	-0.916 (0.286)	0.163 (0.006)	0.286 (0.099)	-0.123 (.099)
Hedmark	-1.571 (0.078)	-2.783 (0.421)	0.172 (0.011)	0.058 (0.023)	0.113* (0.026)
Oppland	-1.419 (0.075)	-1.946 (0.296)	0.195 (0.012)	0.125 (0.032)	0.070 (0.034)
Buskerud	-1.905 (0.083)	-2.944 (0.725)	0.130 (0.009)	0.050 (0.034)	0.080* (0.036)
Vestfold	-1.887 (0.084)	-2.803 (0.728)	0.131 (0.010)	0.057 (0.039)	0.074 (0.040)
Telemark	-1.910 (0.089)	-2.303 (0.655)	0.129 (0.010)	0.090 (0.050)	0.038 (0.051)
Aust-Agder	-1.638 (0.110)	-2.197 (0.609)	0.163 (0.015)	0.100 (0.055)	0.063 (0.057)

Vest-Agder	-1.759 (0.094)	-1.674 (0.445)	0.147 (0.012)	0.158 (0.059)	-0.011 (0.060)
Rogaland	-2.220 (0.079)	-2.451 (0.330)	0.098 (0.007)	0.079 (0.024)	-0.019 (0.025)
Hordaland and Bergen	2.286 (0.066)	-2.923 (0.459)	0.092 (0.006)	0.051 (0.222)	0.041 (0.023)
Sogn Og Fjordane	-1.390 (0.096)	-1.909 (0.309)	0.199 (0.015)	0.129 (0.348)	0.070 (0.038)
More Og Romsdal	-1.791 (0.070)	-3.206 (0.416)	0.142 (0.009)	0.039 (0.016)	0.104* (0.018)
Sor-Trondelag	-1.937 (0.076)	-2.497 (0.393)	0.126 (0.008)	0.076 (0.028)	0.050 (0.029)
Nord-Trondelag	-1.222 (0.083)	-1.756 (0.248)	0.228 (0.015)	0.147 (0.031)	0.080* (0.035)
Nordland	-1.373 (0.059)	-2.203 (0.242)	0.202 (0.010)	0.099 (0.022)	0.103* (0.024)
Troms	-1.367 (0.079)	-2.741 (0.365)	0.203 (0.013)	0.061 (0.021)	0.142* (0.024)
Finnmark	-1.350 (0.100)	-2.639 (0.423)	0.206 (0.016)	0.067 (0.026)	0.139* (0.031)

Source: Norwegian population register

TABLE IV.3. - PARAMETERS ESTIMATED WITH MULTILEVEL MODEL (STANDARD DEVIATIONS IN PARENTHESES)

Characteristic	Multilevel model	
Fixed:	simple	contextual
a_0 (non-farmer)	-1.710 (0.072)	-2.066 (0.111)
a_1 (farmer)	-2.293 (0.137)	-2.003 (0.298)
a_2 (proportion of farmers)		5.400 (1.461)
a_3 (farmers x prop. of farmers)		-8.808 (2.963)
Random:		
$\sigma^2_{u_0}$ (non-farmer)	0.093 (0.033)	0.053 (0.019)
$\sigma_{u_{01}}$ (covariance)	0.057 (0.045)	0.092 (0.040)
$\sigma^2_{u_1}$ (farmer)	0.191 (0.112)	0.214 (0.118)

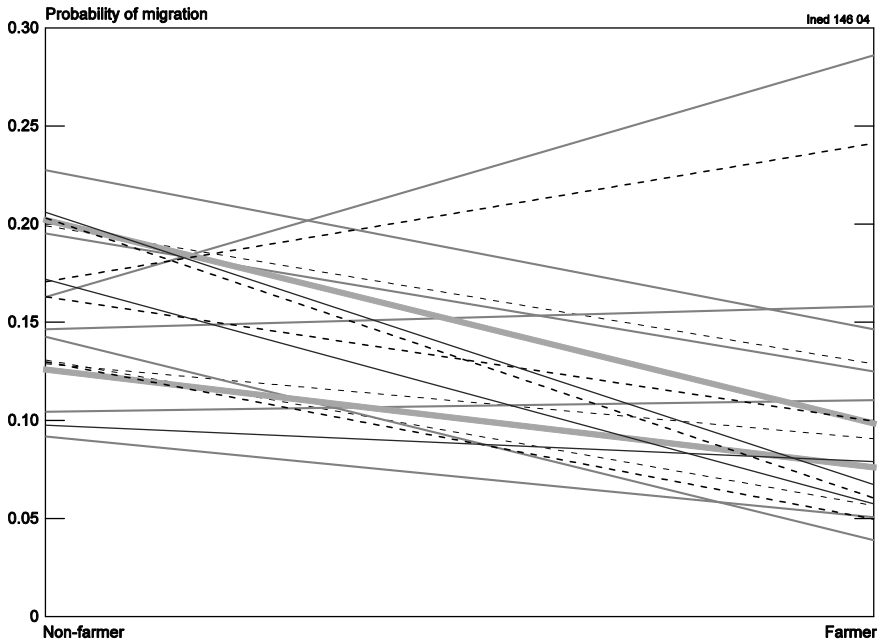


Figure IV.4. - Farmers' and non-farmers' probabilities of migrating estimated separately for each Norwegian region

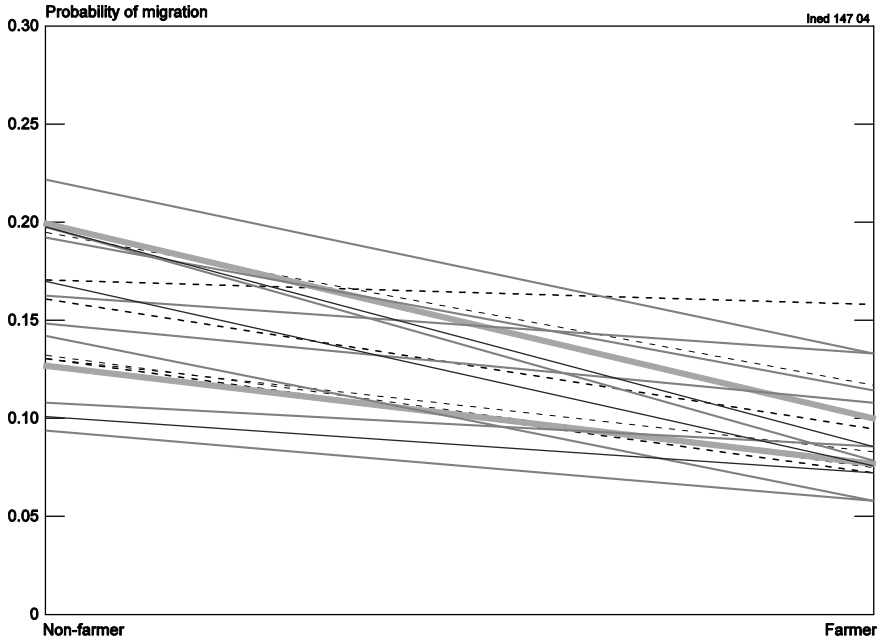


Figure IV.5. - Probabilities of migrating estimated with a simple multilevel model for each Norwegian region

This figure is now closer to figure IV.4—which was based on raw observations—than figure IV.3 was. The dispersions of non-farmer emigration rates are close to the observed values. However, the dispersion for farmers is still too weak, although it comes closer to that of the observations. Furthermore, whereas in figure IV.4 farmers had a higher probability of migrating than non-farmers in four regions (Ostfold, Akerhus, Oslo, Vest-Agder), in this model all regions display a lower probability of migrating for farmers.

Let us therefore see what happens if we incorporate aggregate characteristics into the same model, which we can call a contextual multilevel model, in column 3 of table IV.3, which uses equation IV.14. We see that the estimated parameters remain closer to those of the previous contextual model (table IV.1) and still exhibit a greater dispersion. But, most significantly, we see that the variance for the non-farmers' random variable is almost halved from 0.093 to 0.053. The introduction of aggregate characteristics therefore offers a good explanation of this reduction. At this point it is interesting to see the results estimated with this model (figure IV.6), which we can still compare with the models that gave separate estimates of the parameters for each Norwegian region (figure IV.4).

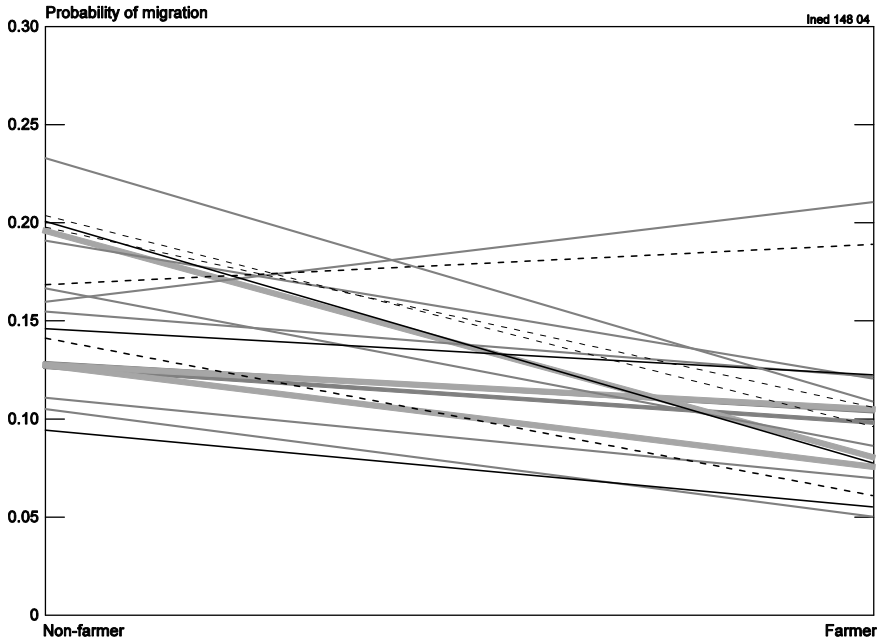


Figure IV.6. - Probabilities of migrating estimated with a contextual multilevel model for each Norwegian region

This figure shows a greater dispersion for farmers than that of figure IV.5: there are now two regions where farmers are more likely to migrate than the rest of the population (Akerhus and Oslo). However, the probability remains slightly lower than the value shown in figure IV.3.

The shift from one model to the next thus gives us a far better match between the data estimated by the model and the data actually observed.

II. - TOWARD A SYNTHETIC PARADIGM

This new approach will not upset the conceptual framework used for event-history analysis, for it continues to operate at the individual level. But, by introducing more complex group effects or spatial divisions relevant to individual behaviors, it will enable us to flesh out the analysis.

The new paradigm will therefore continue to regard a person's behavior as dependent on his or her past history, viewed in its full complexity, but it will be necessary to add that this behavior can also depend on external constraints on the individual, whether he or she is aware of them or not. The contact circle—composed of family members, friends, colleagues or leisure acquaintances—will influence behavior. Likewise, people's environment and the information on the world that they receive from the press and television can play a role in their future actions. More generally, pressure from surrounding society can influence people's behavior without their being fully conscious of it. Persons living in an environment where, for example, unemployment is high may be more willing to accept long-distance migration than if they were living in a full-employment area. Individuals may, of course, be fully aware of the constraints imposed by the society in which they live, and may act in response to them, whether to resist, circumvent or use them if the constraints can improve their status.

Conversely, we shall be able to examine the "perverse effects" of individual actions whose initial aim was totally different from the result obtained (Boudon, 1977). While individual behaviors can be viewed as rational and directed toward a specific goal, it often happens that many people fail to obtain the expected results. There are two major causes of this failure: (1) other individual actions may counter the expected effects, and (2) expectations concerning those actions may be incorrect. The effects are therefore induced by the person's living environment, and we can once again identify them thanks to a multilevel approach. This exercise is particularly worthwhile if we have data on the goals of a given action and its final outcome.

This approach enables us to work on period data as well as on individual event histories situated in a multiple space. The latter will not only consist of standard physical space, in which we can distinguish regions or cities; it can also be a more social space involving networks of interpersonal relationships, a more economic space composed of businesses, institutions, etc., where people work, or any other functional space. In a later chapter, we shall define in greater detail these various environments that can influence individual behavior; for the moment, we merely want to note their complexity.

It is also important to realize that this paradigm allows us to remove the two types of fallacy indicated previously. The ecological fallacy is eliminated, since aggregate characteristics are no longer regarded as substitutes for individual characteristics, but as characteristics of the sub-population in which individuals live and as external factors that will affect their behavior. At the same time, we eliminate the atomistic fallacy provided that we incorporate correctly into the model the context in which individuals live.

Similarly, as noted earlier, the paradigm enables us to transcend the opposition between holism and methodological individualism. This opposition is no longer warranted insofar as we can examine a large number of aggregation levels at once, including an individual level. The society in which we live is composed of many social, economic, political, religious, educational, and other groups, and a given individual is involved in a number of these groups. It is this multiple involvement that will guide people's actions throughout their life histories. In consequence, these effects are what multilevel analysis must identify and examine with the aid of the characteristics operating at each level.

In the second part of this work, we shall need, of course, to take a detailed look at alternative methods for conducting such a multilevel analysis, which brings into play simultaneously the data corresponding to these different observation levels. We shall also have to examine the gaps that persist in this approach—and the means to fill them.

PART II
Multilevel analysis

CHAPTER V

DEFINING LEVELS

Our first task in this chapter is to take a closer look at the many types of groupings of individuals found in all human societies. The diversity of these groups, the hierarchical or more complex links that may exist among them, and their relevance to the study of a human phenomenon justify a fuller description and discussion of them.

We begin by examining the various types of groups that can be studied in social science. We shall try to define them precisely, to see their potential fields of application, and to specify their relevance with respect to other groupings that are harder to construct. Some of these groupings may be considered important in the study of social phenomena, but may actually be difficult to incorporate both into the surveys that would allow their study and into the multilevel analysis that we could perform on them. As a result, we sometimes need to replace them with more conventional groupings that, despite their less clearly visible significance, offer a reasonable proxy.

Next, we look at the potential links between the levels identified in the first part of this chapter. Are the links purely hierarchical or should we assume that they are far more complex? On the face of it, there are grounds for regarding the links as essentially hierarchical: the individual resides in a neighborhood, which is situated in a town, which is located in a county, etc. But many other examples—such as taking into account the neighborhood in which a person lives and the company for which (s)he works—show that alternative, non-hierarchical configurations exist as well. We shall therefore need to review these different types of configurations and classifications and begin to consider the best way of modeling them.

Lastly, we shall have to examine more complex cases—for instance, a situation where the individual level is not the first to be explored but is already an aggregation of earlier levels. This can happen, for example, when the stages of an individual life are regarded as the first levels and the individual herself or himself forms the second level. The situation also becomes more complicated when we want to perform a multilevel event-history analysis. There is no reason why an individual should remain in the same entity of a given aggregation level during the entire observation. We therefore need to take account of such situations and develop plans to enable them to play their proper role in the model.

I- Different types of levels

We cannot describe here all the types of levels that can be examined in social science, but only those most often used, particularly in demography and related disciplines. We begin with those that seem, in principle, the most "natural" choices: groupings of units that are easy to define and to distinguish from one another, and whose effect on behaviors seems entirely possible. Next, we look at other levels that have a less manifest effect on behaviors or, on the contrary, that are harder to define and are less easily distinguished from one another.

Social or economic groups with obvious effects

Among all these groupings, the first that comes to mind is the family: it results from the links involved in the reproduction process, especially as these links are sanctioned by legal provisions or customs (Henry, 1981). We shall not go into the details of more precise definitions that take account of these laws and customs, which vary from one country or ethnic group to another. But we can see that a family is a complex group where parents and children play very different and even conflicting roles. This dissymmetry of roles partly undermines the value of the family in the aggregate for multilevel analysis, in which we are looking for what unites group members rather than what divides them. For our purposes, it may be useful to separate the family into two groups: parents and children. By doing so, we can see the value of working on the children group to study phenomena such as the age when children leave their parents (Murphy and Wang, 1998), or, alternatively, working on the parent group to study the types of successive jobs (Courgeau and Meron, 1996).

The concept of household—i.e., the set of persons usually living together under the same roof (Henry, 1981)—is even more complex than the family as a whole. Again, if we want to use the household concept in a multilevel analysis, it will generally be useful to decompose the group into more homogeneous sub-groups. The same applies to the concept of contact circle (Lelièvre et al., 1997, 1998), which combines members of the different households to which the individual has belonged during his or her lifetime and key family members or close friends, who are not necessarily coresident. This notion is very important for a better assessment of the effect of adjacent or competing processes that will influence an individual's event history. One way to model the contact circle is to treat each group as a composite individual, described not only by exogenous collective variables, but by the set of characteristics of group members. These models may be called pseudo-individual. The alternative, multi-individual modeling is much harder to implement, and tests of multivariate modeling of such sets are still in their initial phases (Lelièvre et al., 1997, 1998).

The grouping of dwellings intended to accommodate a household in apartment buildings will give us a first example of hierarchically arranged levels: the individual at the first level, the household at the second, and the building at the third. In certain cases, when the individual lives in a private house, this third level will comprise only one household present in the building. This segmentation is relevant insofar as we know that households living in the same building are much closer to one another than households living in different buildings or living in a house and a building.

Other groupings, of an economic nature, can have a major effect on their members' behavior. The business firms or organizations in which individuals work are totally relevant to the analysis, as they can influence many behaviors of their members. Again, the complexity of the structure of a firm or organization may lead the social scientist to include only one of its

divisions in the study. It is often useful to take into account the members who are close to the individual examined—for example, to restrict the study to persons working in the same workshop or belonging to the same socio-occupational category. As membership in the same entity can bring people into frequent contact, this can induce similar behaviors in many areas connected with the workplace, the family or politics.

Other communities are important for younger people of school or student age. First, the class to which they are assigned can strongly influence their scholastic performance and, more generally, their behaviors in other aspects of their lives. The teacher's effect can be overriding, and we must detect it accurately. Social-science multilevel analysis is, in fact, a tool of education science, for which it was essential to demonstrate this effect and its importance. Researchers used such an approach to examine the effect of teaching methods that produced widely differing results among students when the analysis ignored their grouping under a given teacher (Bennett, 1976). The analysis showed that the effect disappeared when the grouping was properly specified in a multilevel model (Aitkin et al., 1981).

For the young, we should also consider another entity at a more aggregated level than the class: the school or university. Attendance at a particular school or university is not neutral with regard to a student's performance and other aspects of his or her life. As with buildings, if we want to obtain meaningful results, it is preferable to use a hierarchical model comprising three aggregation levels: the individual at the first, the class at the second, and the school at the third. We could add a fourth level to represent the education district, but its value seems more limited owing to the diversity of institutions located in a given district.

The analysis of health and mortality will require other types of segmentations. The first type is a distinction between hospitals and clinics, in order to compare their results in terms of healing and mortality. Again, it will be useful to distinguish between the departments of these hospitals and clinics by specialty, so as to better compare like with like. Many studies in epidemiology make this distinction, revealing sometimes sizable effects (Matsuyama et al., 1998; Diez Roux, 2003). A second type of segmentation may concern patients of the same doctor, such as a general practitioner, to compare the prescriptions given to persons suffering from the same illness or displaying the same symptoms.

We must also consider groupings of individuals into cities or metropolises—in human geography—whose functioning and high stability over time prove their relevance as division units. A city is a well-delimited entity whose self-evident economic role has been demonstrated by geographers and economists. Specifying it as an aggregation level in a multilevel analysis can yield new results that we could not obtain if we worked only on its inhabitants or, on the contrary, on cities in the aggregate. We can also classify cities and towns by size category so as to examine a smaller number of basic units; the categories reflect a ranking of population thresholds that trigger the presence of certain kinds of infrastructure and facilities (Pumain and Saint-Julien, 1990).

Another type of observation will lead us to regard the individual as located not at the first aggregation level but at a higher level: the multistage survey. In this kind of survey, the first level of observation will be the stage in which values are measured; the second level will be the individual observed in each stage. As we can see, there is usually a wider variation between individuals than between the values measured for a single individual. For instance, if we measure a woman's completed fertility, its value will vary in a much narrower range between successive stages than the variation among women in the sample.

We could easily continue this review of social groups that exert manifest effects on the behavior of their members: sports teams, leisure groups, congregations, membership organizations, and so on. We prefer to turn now to a more detailed examination of other groupings that seem more external to individuals' actions but often emerge as unquestionably significant in multilevel analyses.

Geographic and administrative groups with less direct effects

Let us start with administrative groupings, which divide a country into units that—on the face of it—have little to do with their inhabitants' behaviors. In France, for example, the division into municipalities (communes), départements and "program regions" (i.e., regions with separate local-development programs) forms a system of nested units. The first two categories date, for the most part, from 1790, but their origin is much older (Pumain and Saint-Julien, 1990). The French commune is a venerable unit of rural life that revived and enlarged the scattering of Gallo-Roman villae during the Christianization of the fourth and fifth centuries. The national territory is apportioned among the 36,000 or so municipalities: each administers an average of only 15 square kilometers and 1,500 inhabitants, the "main towns" (chefs-lieux) being almost always less than 5 kilometers apart. The chef-lieu of the département is located at the center of the district (circonscription), so that local residents could generally travel there on foot in a day. The 83 initial départements, which became 90 until 1968, have increased to 95, owing to the very dense urbanization of the Paris region. The division into 22 "program regions" is more recent: its gradual implementation, driven by territorial-management requirements, began in the late 1960s. Similar but often less detailed segmentations exist in all countries for administrative purposes.

As we can see, this administrative segmentation, established long ago and relevant at the time, now seems less linked to individual behaviors than the groups discussed earlier. Of course, if we want to study inter-regional migrations, we will need to adopt the regional segmentation (Courgeau, 2003c). But many other studies, as well, are performed with these administrative divisions because data are often collected and published at their level, in particular by most statistical agencies: this includes data on population, mortality, fertility, nuptiality, health, education, the economy, and so on. Now there is no obvious reason why administrative divisions should influence these phenomena. Nevertheless, the multilevel analyses that incorporate them show that the divisions often yield unquestionably significant results. We must therefore assume that they constitute usable proxies of other divisions, better suited to these studies, but for which no statistics are gathered. This should not prevent us, however, from seeking to identify more clearly the relevant divisions that the current statistics merely proxy. For instance, the rural town proxies an individual's network of relationships, which could be clearly identified only through more in-depth surveys (Courgeau, 1973).

Other types of segmentations are used by survey statisticians to reduce data-collection costs. These techniques have been developed to supply valid statistical inferences at lower cost, for example to compare mean values or fit regression models. They sometimes use administrative divisions, but often geographic units or other segmentations that are more satisfactory for the study to be performed. They have rarely been studied in their own right, for they were a means to improve estimates and not an end in themselves. However, a multilevel approach can find new value in this structure, which is no longer merely an observed sample but also a means of revealing the structure of the population studied.

At a higher aggregation level, we can examine groupings of countries. The effect of national policies on the behaviors of these countries' inhabitants is obvious and makes this

segmentation worth using. It has been employed, for example, by Wong and Mason (1985) to study the use of contraceptive methods in a number of developing countries.

II Different types of nestings

After examining the wide variety of types of levels that can be studied, we must now see how these levels are organized in relation to one another. We shall need to examine two different cases, depending on whether the observations take place at a given moment (period or cross-sectional observation) or, on the contrary, over time or during the life of the individuals observed (longitudinal or event-history observation).

Period observation

The simplest and most often observed organization is a hierarchical classification. For instance, individuals (level-1 units) are situated in municipalities (level-2 units), which, in turn, belong to départements (level-3 units), which are grouped together into regions (level-4 units). Each segmentation fills the entire national space, and at the same time a higher-level segmentation comprises a certain number of lower-level units: a département is a set of municipalities and a region is a set of départements. Naturally, the sample observed cannot be exhaustive, like a census. However, the observation requires some precautions. Let us assume that we want to observe all regions, but that for cost reasons we cannot observe all départements and—even less—all municipalities. We therefore need to start by selecting a sample of départements, then in each département a sample of municipalities, so as to respect the hierarchy.

We have already noted other types of hierarchical segmentations: students grouped together in classes, which are grouped together in schools, which belong to education districts. Likewise, we could place the individual at the first aggregation level, the household at the second, the building or neighborhood at the third, and so on.

We can schematize these different levels as in diagram 1: the first level is the individual, the second is the municipality, and the third is the département. The first four individuals are in one municipality, the next five in a second, and the following in a third: all three municipalities belong to the same département.

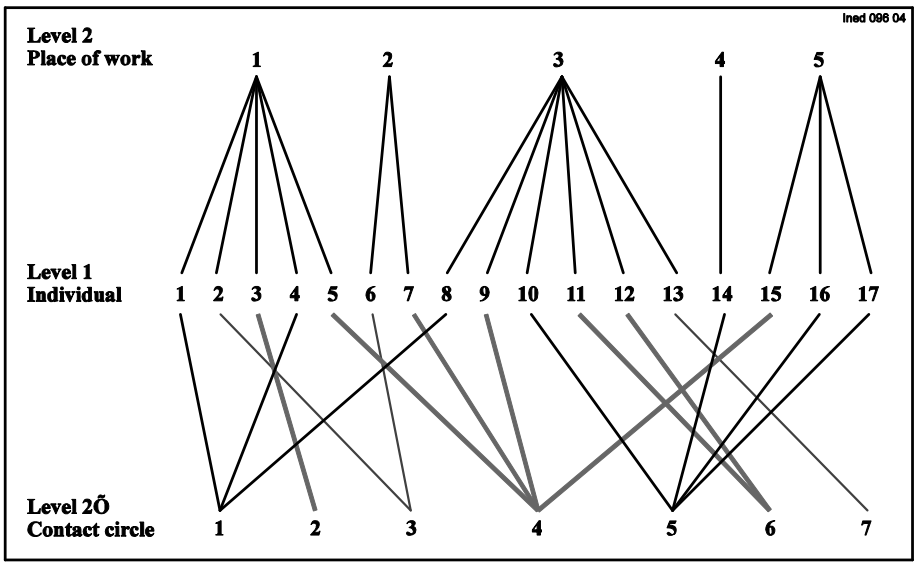


Diagram 1. - Hierarchical classification

Inter-level relationships can, however, be more complex. Cross-classifications allow us to work with levels that have no reason to be hierarchized. For instance, a given individual is simultaneously involved in his or her family environment—represented, for example, by his or her contact circle—and in his or her workplace. There is no reason to regard these two environments as hierarchized, despite the perfectly natural assumption that they may have a separate and joint effect on many individual behaviors. Multilevel models must therefore be capable of taking such classifications into account.

Diagram 2 depicts a cross-classification. We see, for example, that individuals 1 and 4 have the same workplace and the same contact circle; individual 2, instead, has the same workplace and 1 and 4 but a different contact circle; individual 6 has the same contact circle as 2 but a different workplace, and so on.

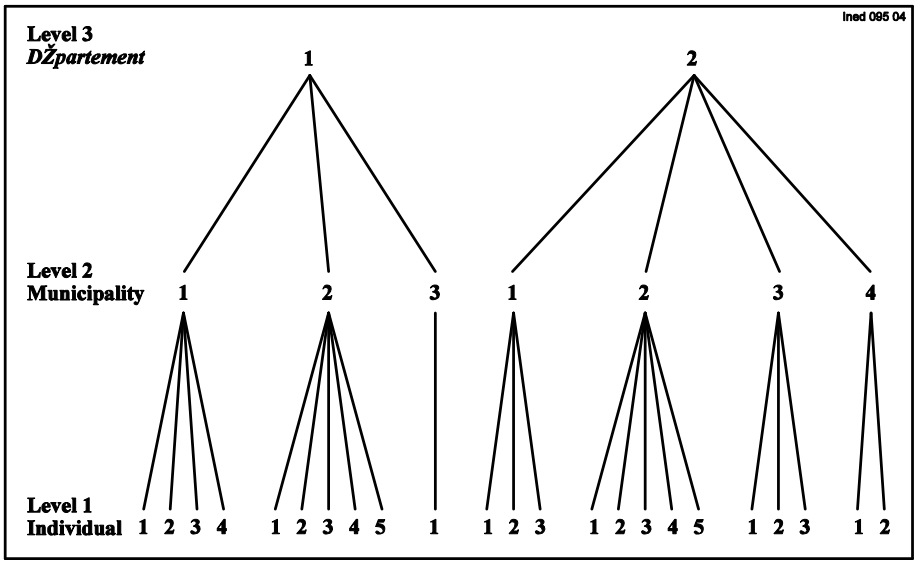


Diagram 2. - Cross-classification

We can also find combinations of hierarchical classifications and cross-classifications involving different levels used in an analysis. For instance, to the previous cross-classification between family and work, we can add a hierarchical classification between family-residence locations, which can be situated in municipalities, départements, and regions.

Lastly, there are classifications in which individuals can belong to several units of a given higher level simultaneously. In diagram 2, the same individual may belong to several contact circles—for example to one as a family member and to another as close friend—or to several workplaces, when he or she holds multiple jobs. Such cases can be described as multiple participation models.

Event-history observation

The previous diagrams were based on period (i.e., cross-sectional) observation, in which we studied the occurrence of no more than one event per individual. If we now incorporate time or length of life, the situation will grow more complicated: several events can occur for the same individual, and his or her position in different aggregation levels can change. Let us illustrate this with a specific example.

Suppose we want to study the fertility of individuals by region of residence. We know that, on a period basis, fertility varies significantly from one region of residence to another. In France, the peak values form a crescent spanning the northern, western, and eastern regions; the levels have varied over time, but within fairly stable geographic boundaries. If we now want to observe fertility changes longitudinally in a given generation, the task becomes more complex, and we shall need additional hypotheses to perform the analysis. This is because many individuals will migrate during their fertile life, and we shall need to factor in the possible changes in behavior determined by their place of residence.

The first hypothesis is to assume that migration causes an instant change in fertile behavior. If so, we can plot individual situations on a diagram that generalizes diagram 2 above, in which the first level records events (migrations as well as births) occurring over time, and in which two non-hierarchized levels will coexist. One of these levels will be linked to the

individual, the other to the observed regions in which individuals have lived. Diagram 3 depicts this situation. The first individual has a first child in region 1, then migrates to region 3, where (s)he has a second child. The second individual migrates from region 2 to region 3 without yet having a child, then has his or her two children in region 3. The third individual has a first child in region 2, then migrates to region 4 where (s)he has a second child, and eventually migrates to region 5 where (s)he has a last child.

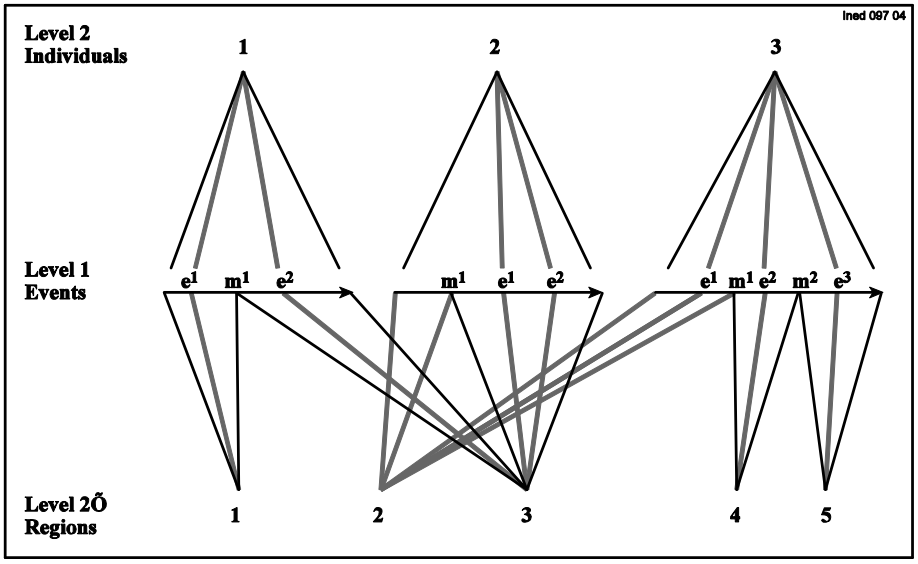


Diagram 3. - Multilevel event-history model

This hypothesis seems too extreme, however, and individuals may adopt the behavior of the destination region over a time interval in which their fertility may shift from that of their region of origin to that of their region of arrival. This interval is unfortunately not known and may vary from one woman to another. We shall need to set up a far more complex model, using new hypotheses that will be tested with the aid of observations.

Lastly, if we want to analyze the migrations themselves, restricting the explanatory variables to the characteristics of the places of departure does not seem a very satisfactory option. We need to incorporate the information available to individuals on possible destinations and their attractiveness for these potential migrants. The analysis of this information must also include the individual relationship networks, which will be able to supply that information. Once again, therefore, we are dealing with a very complex research topic that exceeds the scope of multilevel analysis itself, a tool we need to elaborate in order to carry out this study.

III- Conclusion

At the end of this examination of aggregation levels and their reciprocal relationships, we can note their large number and the complexity of their potential interactions, even though we have given only a partial description of them here. We must also stress the fact that we are still a long way from having identified all the levels. Considerable research remains to be done to determine the most relevant levels for understanding our society more fully. We need to investigate and analyze the networks of interpersonal relationships in all their complexity. It is

not enough to pick a sample of individuals and question them about their relationship network. We also need to clearly identify the structure of interlocking networks that exist in a society (Degenne and Forsé, 1994,1999).

In this conclusion, we must also stress the fact that the oversimplified relationships that we have identified between these levels are much more complex and need to be studied at the same time as we introduce them into multilevel analyses. An individual's involvement in several units of a given level will create problems due to often arbitrary weighting, which we shall need to examine in greater detail and perhaps solve in a different way. Likewise, the introduction of event histories viewed in a multilevel perspective requires a much deeper examination of the methods used and their application. An event history takes place in a complex space that changes continuously in the course of a lifetime. To take account of this complex space, we must generalize multilevel models with multiple memberships. Similarly, the models that incorporate time are still too simple to take temporality changes into proper account.

CHAPTER VI

LINEAR ANALYSIS OF CONTINUOUS CHARACTERISTICS

Continuous individual characteristics are rather infrequent in demography. Demographers very often use binary characteristics: an individual is alive or deceased, married or not, has a first child or not, has migrated for the n^{th} time or not, and so on. Polytomous characteristics can also be used: an individual is married, never-married, widowed, divorced or remarried, for example. Demographers can also study the occurrence of an event during an individual's lifetime: death, first marriage, first child's birth, etc. We can easily see that the analysis of such cases requires more complex models—of the logit, polytomous or event-history type—than a regression on continuous characteristics. Also, by adopting a retrospective viewpoint on an individual's life and observing cumulative events, the demographer will generally be unable to regard them as continuous characteristics. Even when the number of observed events is large—such as the number of migrations performed since a person began living on his or her own—it is often preferable to regard the number as an event count and to use log-linear models to analyze it.

However, other social sciences use continuous characteristics more often. Two examples are education science, when it examines students' grades at different moments in their schooling, and economics, when it studies a person's wages or income at different dates. In fact, demographers often participate in such studies, in particular because of the demographic characteristics incorporated into the analysis.

For these reasons, it is useful to begin our examination of the different types of multilevel models by that of linear-regression models. These models have long been used in the social sciences and are the simplest to estimate, the others requiring approximations that make them harder to calculate. We should, however, have a clear perception of their underlying hypotheses and the conditions in which to apply them.

We begin with the simplest case of a two-level linear-regression model. First, we describe the methods for estimating the model's parameters, their variance, and the residuals, which should be examined. We apply these methods to a demographic example in order to shed light on their results, and we compare the results with those of an analysis at the aggregated level. We can then move on to a more general model incorporating a larger number of levels.

I. A two-level linear-regression model

Let us first suppose that we want to study the effect of a continuous characteristic of individual *i* situated in an area *j*, x_{1ij}, on another, equally continuous characteristic, y_{1ij} The number of individuals present in area *j* is n_j and the total number of areas is *J*. If we work on each area, we shall be able to estimate a corresponding number of linear regressions of the following type:

$$y_{ij} = a_{0j} + a_{1j}x_{1ij} + \varepsilon_{ij} \qquad \text{(VI.1)}$$

where the parameters a_{0j} and a_{1j} will depend on the area *j* and where the residual ε_{ij} is supposed to follow a Normal distribution N(0, σ_j). If the number of areas is large, as some of them may have a small number of respondents, these estimations may yield weakly significant results when the background noise prevails over the main phenomenon we want to analyze: the link between individual characteristics, modulated by the area where the individual is located.

Hence the notion of using a single model for all areas, but a model whose parameters will depend on a random variable. For this purpose, let us write:

$$a_{0j} = a_0 + u_{0j} \qquad \text{and} \qquad a_{1j} = a_1 + u_{1j} \qquad \text{(VI.2)}$$

where u_{0j} and u_{1j} are zero-mean random variables whose variances and covariance are:

$$\text{var}(u_{0j}) = \sigma_{u0}^2 \qquad \text{var}(u_{1j}) = \sigma_{u1}^2 \qquad \text{and} \qquad \text{cov}(u_{0j}, u_{1j}) = \sigma_{u01} \qquad \text{(VI.3)}$$

The overall model can thus be written:

$$y_{ij} = a_0 + a_1x_{1ij} + (u_{0j} + u_{1j}x_{1ij} + \varepsilon_{0ij}) \qquad \text{(VI.4)}$$

where the variance of ε_{0ij} is equal to:

$$\text{var}(\varepsilon_{0ij}) = \sigma_{\varepsilon 0}^2 \qquad \text{(VI.5)}$$

and the variance of the level-2 random variable is equal to:

$$\text{var}(u_{0j} + u_{1j}x_{1ij}) = \sigma_{u0}^2 + 2\sigma_{u01}x_{1ij} + \sigma_{u1}^2x_{1ij}^2 \qquad \text{(VI.6)}$$

The model therefore comprises a fixed part and a random part in parentheses in equation (VI.4), and requires the estimation of only six parameters versus $3\times J$ parameters for model (VI.1). These random variables are still called residuals, as in single-level models. However, they are now situated at two aggregation levels. We therefore set an underlying hypothesis of independence between random variables at the individual level ε_{0ij} and random variables at the aggregate level u_{0j} and u_{1j}, whereas a dependence exists between the latter two random variables. We can easily see that this reduces the number of parameters we need to estimate, and this can also have a very favorable impact on their significance. But it is important to realize that we cannot achieve this simplification unless the previous hypothesis is confirmed.

The model also enables us to estimate the residual terms a_{0j} and a_{1j} for each area, as we shall see later. We can thus identify the areas with the most extreme results, for which we

can test the divergence from the mean or from results in another area. We can then perform a more detailed study on these areas to identify the reasons for their divergence.

It is easy to generalize the model so as to bring a large number of characteristics into play, x_{kij} being the k^{th} of the m characteristics introduced, with random variables at the aggregate level corresponding either to each of these characteristics or only to some of them. The model (VI.4) can accordingly be written:

$$y_{ij} = a_0 + \sum_{k=1}^{m} a_k x_{kij} + (u_{0j} + \sum_{k=1}^{m_1} u_{kj} x_{kij} + \varepsilon_{0ij}) \qquad \text{(VI.7)}$$

where u_{kj} is the random variable corresponding to the k^{th} characteristic. Naturally, we can restrict the sum to selected characteristics, for their effect may be insignificant or even null: if so, $m_1 < m$. We must then consider the $\frac{m_1(m_1+1)}{2}$ level-2 variances and covariances to estimate in addition to the $m+1$ parameters.

We can also generalize these models by adding more levels. For example, when we have the results of a survey based on an area sample, such as the French labor-force survey, we can use the different degrees of stratification as levels of analysis. The formalization of these models resembles that of a two-level analysis, which we have just described.

II. Estimating the model parameters

This type of model, which comprises random variables situated at different aggregation levels, requires more complex estimation methods than those usually employed to estimate the parameters of a traditional linear-regression model, such as the least-squares method.

Statisticians have gradually developed these methods (Hartley and Rao, 1967; Harville, 1976; Miller, 1977; Aitkin et al., 1981; Mason et al., 1984; Goldstein, 1986) by making them applicable to increasingly complex situations. The methods generally operate by successive approximations to converge toward acceptable estimates, for analytical solutions such as those used for standard regressions are no longer applicable.

Rather than elaborating on the methods here, we shall describe their general principle, in the simple case consisting of an explanatory variable, x_{1ij}, a regional random variable, u_{0j}, and an individual random variable, ε_{0ij}, in equation (VI.4), which becomes:

$$y_{ij} = a_0 + a_1 x_{1ij} + (u_{0j} + \varepsilon_{ij}) \qquad \text{(VI.8)}$$

The principle of this method consists in switching from a estimator where we assume we know the variance of u_{0j} (enabling us to estimate the values of the parameters a_0 and a_1 using the generalized least-squares method) to another estimator, which uses the values of the parameters we have just estimated and allows us to improve the estimate of the variance of u_{0j}. Let us see in greater detail how to proceed.

To begin, let us assume that the variance of u_{0j} is zero. Using the ordinary least-squares method, we can estimate the parameters a_0^1 and a_1^1, which minimize the variance of e_{0ij}:

$$\hat{a}_1^1 = \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j} y_{ij}x_{1ij} - \frac{1}{\sum_{j=1}^{J} n_j}(\sum_{j=1}^{J}\sum_{i=1}^{n_j} y_{ij})(\sum_{j=1}^{J}\sum_{i=1}^{n_j} x_{1ij})}{\sum_{j=1}^{J}\sum_{i=1}^{n_j} x_{1ij}^2 - \frac{1}{\sum_{j=1}^{J} n_j}(\sum_{j=1}^{J}\sum_{i=1}^{n_j} x_{1ij})^2} \qquad \text{(VI.9)}$$

$$\hat{a}_0^1 = \frac{1}{\sum_{j=1}^{J} n_j}\sum_{j=1}^{J}\sum_{i=1}^{n_j} y_{ij} - \frac{\hat{a}_1^1}{\sum_{j=1}^{J} n_j}\sum_{j=1}^{J}\sum_{i=1}^{n_j} x_{1ij} \qquad \text{(VI.10)}$$

We can accordingly calculate the residuals for this estimation, $\hat{r}_{ij}$

$$\hat{r}_{ij} = y_{ij} - \hat{a}_0^1 - \hat{a}_1^1 x_{ij} \qquad \text{(VI.11)}$$

Having determined these residuals, we can improve the estimation of the variances of u_{0j} and e_{0ij}, which are, it will be recalled, independent random variables. If we represent the residuals as a column vector R, we see that the mathematical mean of the product of this vector by its transposition, the line vector R^T, is the variance-covariance matrix for the two observation levels (in this instance, the covariances are null). The terms of this matrix are equal to $\sigma_{u0}^2 + \sigma_{e0}^2$ for a single individual, to σ_{u0}^2 for two individuals present in the same area, and to zero when these two individuals are in different areas.

To simplify the computation, we can rewrite this matrix as a column vector by eliminating the null terms that contain no information and arranging the columns in sequence. We obtain a vector with $\sum_{j=1}^{J} n_j^2$ terms, and the relationship for estimating σ_{u0}^2 and σ_{e0}^2 can be written:

$$\begin{pmatrix} r_{11}^2 \\ r_{11}r_{21} \\ r_{11}r_{31} \\ \vdots \\ r_{11}r_{n_{11}} \\ r_{21}r_{11} \\ r_{21}^2 \\ r_{21}r_{31} \\ \vdots \\ \vdots \\ r_{n_J J}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 \\ \sigma_{u0}^2 \\ \sigma_{u0}^2 \\ \vdots \\ \sigma_{u0}^2 \\ \sigma_{u0}^2 \\ \sigma_{u0}^2 + \sigma_{e0}^2 \\ \sigma_{e0}^2 \\ \vdots \\ \vdots \\ \sigma_{u0}^2 + \sigma_{e0}^2 \end{pmatrix} + R' = \sigma_{u0}^2 \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} + \sigma_{e0}^2 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} + R' \qquad \text{(VI.12)}$$

where R' is a residual vector. This produces a new system of linear equations where the parameters to be estimated now consist of the variances σ_{u0}^2 and σ_{e0}^2, all of whose variables are known, and whose residual R' must be minimized. The generalized least-squares method enables us to estimate these parameters, under the hypothesis of a Normal distribution. Even if this hypothesis is not confirmed, the parameters will be estimated correctly, but not their

standard deviation. To solve this difficulty, we must use other distributions or estimation methods (for more details, see Goldstein, 2003).

As we perform further iterations, we must use these estimated variances to find better estimates of the parameters a_1 and a_0. The generalized least-squares method gives us those estimates, $\hat{a}_1^2$ and $\hat{a}_0^2$. We continue the iterations until the procedure converges, i.e., until all the parameter and variance estimators stay nearly constant from one iteration to the next. For this purpose, we set a convergence limit.

In a classic model, it is easy to determine the regression's residual for each individual as the value of $\hat{r}_{ij}$. But in a multilevel model, we shall also have at our disposal the residuals at the regional level. Let us look briefly at how to estimate u_{0j}, for example. We can write:

$$\hat{u}_{0j} = E(u_{0j} \mid r_{ij}, \hat{V}) \qquad \text{(VI.13)}$$

where $\hat{V}$ is the matrix of the estimated variances-covariances. We can show that the solution to this equation yields the following estimator of $\hat{u}_{0j}$:

$$\hat{u}_{0j} = \frac{\sigma_{u0}^2 \left(\sum_{i=1}^{n_j} r_{ij} \right)}{\left(n_j \sigma_{u0}^2 + \sigma_{e0}^2 \right)} \qquad \text{(VI.14)}$$

We shall not elaborate on these estimation methods here; we refer the reader interested in these statistical issues to the publications by Goldstein (1986, 1989) and the detailed presentation in his book, Multilevel Statistical Models (2003). Many software programs now allow the application of these models, and we provide a list of them in the appendix.

We prefer to dwell here on a specific application of the models, which will enable us to see what we can draw from a multilevel linear-regression analysis.

Example no. 6: Wages in survey as a function of initial wage, in France

This example is based on the "Youth and careers" survey conducted by INSEE (the French National Statistical Institute) in 1997. A sample of persons born between 1952 and 1983 were asked about their family life, working career, migration history, and various characteristics of the respondents themselves and their families. For this example, we shall take the respondent's wage at the start of his or her career (first job held for more than six months) and the wage reported at the time of the survey. Persons who did not report a wage for different reasons (self-employed, voluntary non-response, etc.) at the two dates were excluded from the observation sample.

We restrict our analysis to members of generations born between 1952 and 1961, whose mean age at the time of the survey is therefore 40 years. The sample selected on this criterion comprises 4,777 individuals. The multilevel model examines two aggregation levels: the first is the individual, the second the département of residence of the respondent's parents, when the respondent was 15 years old. However, to avoid too small an observation sample in thinly populated départements, we decided to amalgamate the départements with fewer than 200,000 inhabitants (Hautes-Alpes, Ariège, Gers, Haute-Loire, Lot, Lozère, and Territoire de Belfort) with a nearby département in the same region (Alpes-de-Haute-Provence, Haute-Garonne, Tarn, Cantal, Lot-and-Garonne, Gard, and Doubs respectively). We also treat overseas départements and territories (DOM-TOMs) and the rest of the world as a single area. This gives us a segmentation into 89 areas.

The dependent variable is the wage declared in the survey, as a function of the initial wage and selected characteristics of the individual at the start of his or her career: first socio-occupational category, generation, and sex. We included all persons who reported a wage at both dates, irrespective of whether they were working full-time or part-time. To avoid extreme spreads—which may be due, in particular, to misreporting or reflect very special situations—we have eliminated individuals who declared very low wages (under 120 francs for the initial wage and under 500 francs for the final wage) and very high wages (over 15,000 francs for the initial wage and 50,000 francs for the final wage). While this does not greatly alter the effects identified by the models used, it does make them more significant. For the estimations, we divided the wage by 1,000.

At this point, we must address the issue of wage modeling. Classical economic studies assume lognormal wages. This assumption is justified on the grounds that the individual wage is determined by a large number of effects, all of them weak, mutually independent, and operating in a multiplicative manner (Depardieu and Payen, 1986). As a rule, economists will therefore assume that the various individual characteristics affecting wages will exert a multiplicative effect. Economic models will take the log wage as the dependent characteristic, on which the individual characteristics will exert a linear effect: for continuous characteristics (earlier wage, length of working time, and so on), logs will be examined; discrete characteristics such as sex, educational attainment, socio-occupational category, etc. (Depardieu, 1978; Glaude and Jarousse, 1988; Chabanne and Lollivier, 1988) will be treated as n-1 binary characteristics, when the number of their values is n.

However, the detailed analysis of our data, without distinction by département, shows that linear regressions on the characteristics, as well as the examination of the correlations between them, yield far better results when we work on the raw data rather than on their logs. For instance, the model incorporating only the initial wage yields an R^2 of 0.16 with raw values versus 0.08 using logs. Likewise, most correlations between final wage and characteristics are more significant with raw values than with logs. When we estimate a linear-regression model, it is the residuals and not the variables themselves that must be normally distributed for the estimations of the parameters and their variance to be correct. We shall therefore begin by working on raw data. Later on, however, we shall see that the hypotheses of a multilevel regression model are not properly confirmed and we shall try to remedy this problem.

First, we consider the effect of the initial wage and its mean value by region on the final wage, by estimating models of growing complexity; next, we look at the effect of the other characteristics, while examining the hypotheses set for these estimations.

Table VI.1 shows the result of the estimations of these multilevel linear models, restricted to the initial wage and its mean value by region, with the parameters' standard deviations in parentheses.

TABLE VI.1. - EFFECT OF INITIAL WAGE AND SEX ON FINAL WAGE
(STANDARD DEVIATION IN PARENTHESES)

Parameters	Multilevel model			
Fixed	Model 1	Model 2	Model 3	Model 4
a_0 (constant)	9.105 (0.130)	6.946 (0.129)	7.012 (0.139)	5.728 (0.447)
a_1 (initial wage)	-	0.982 (0.034)	0.951 (0.060)	0.942 (0.060)
a_1 (mean initial wage)	-	-	-	0.592 (0.195)
Random, level 2				
σ^2_{u0} (constant)	0.931 (0.220)	0.506 (0.141)	0.629 (0.239)	0.529 (0.221)
σ_{u01} (covariance)	-	-	-0.202 (0.089)	-0.197 (0.087)
σ^2_{u1} (initial wage)	-	-	0.167 (0.045)	0.172 (0.046)
Random, level 1				
σ^2_{e0}	24.21 (0.504)	20.71 (0.427)	20.23 (0.421)	20.22 (0.421)
-2 (log-likelihood)	28272.8	28104.3	28060.5	28051.8

Model 1 gives us the mean wage at 40 years in thousands of francs (i.e., 9,105 francs), with a very strong variance at the individual level (24.2), and sizable differences between regions, resulting in a variance at département level of 0.931, which is entirely significant. Model 2 shows that the initial wage influences the final wage, with a proportionality coefficient near unity (0.98): we can conclude that the final wage is approximately equal to the initial wage plus 7,000 francs. The variance at the département level is almost halved and the individual variance is reduced by nearly one-fifth. It is useful to calculate the proportion of total variance taken into account by the aggregate level:

$$p = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2_{e0}}$$

(VI.15)

Excluding the initial wage, the proportion is equal to 0.037; including the initial wage, the proportion falls to 0.024. The introduction of a département level therefore covers only a small share of total variance. The reduction in value of minus two times the log-likelihood replaces the variation of R^2 in a classic regression model: we can thus verify, by means of a χ^2 test, that the introduction of the initial wage effectively improves the model's quality. But we can no longer calculate the share of variance explained by the model, as with R^2 in a standard regression.

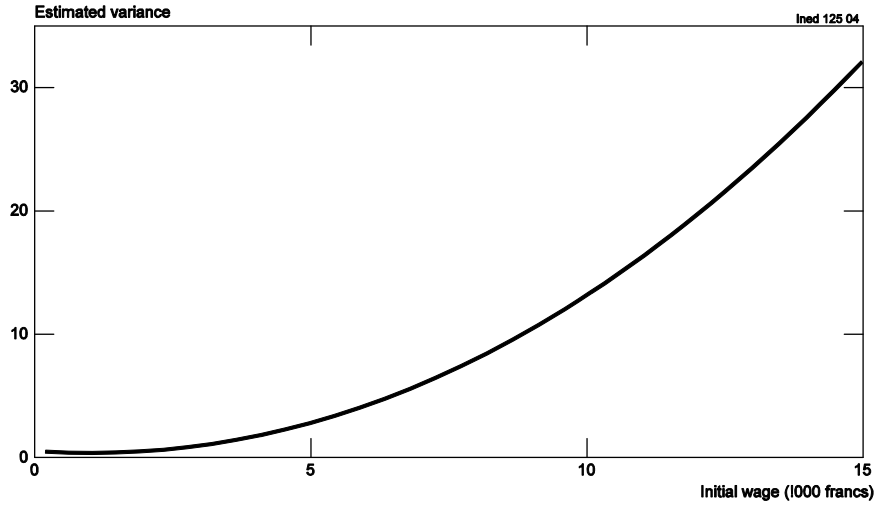


Figure VI.1. - Estimated variance at area level as a function of wages

Model 3 modulates the département variance as a function of the initial wage. This entails few changes in the fixed characteristics but major, significant changes in département variance, which is now a function of the initial wage, and on individual variance, reduced by nearly one-quarter. We can compute the effect of the initial wage département variance:

$$\text{var}(u_0 + u_1 x_{1ij}) = \sigma^2_{u0} + 2\sigma_{u01} x_{1ij} + \sigma^2_{u1} x^2_{1ij} = 0.629 - 0.404 x_{1ij} + 0.167 x^2_{1ij}$$

(VI.16)

Figure VI.1 reports the variance observed at area level as a function of the initial wage. As we can see, the variance is minimal with a value of 0.57 for a very low initial wage on the order of 600 francs, and rises to over 30 for an initial wage of 15,000 francs. It is easy to combine the direct effect of school-leaving age with the variations indicated in figure VI.1: this leads to figure VI.2, which plots the mean final wage as a function of the initial wage, with a 95% confidence interval. We see the increased importance of the département effect on the final wage, when the initial wage rises. Note the correlation between the y-intercept (approximately the final wage when the initial wage is very low) and the slope of the straight lines measuring the départements is negative and equal to –0.62: this denotes the fact that when the y-intercept is low, the slope of the line for a given département is strong; conversely, when the y-intercept is strong, the slope is weak. This relationship does not obtain, however, given that the correlation coefficient does not equal –1.

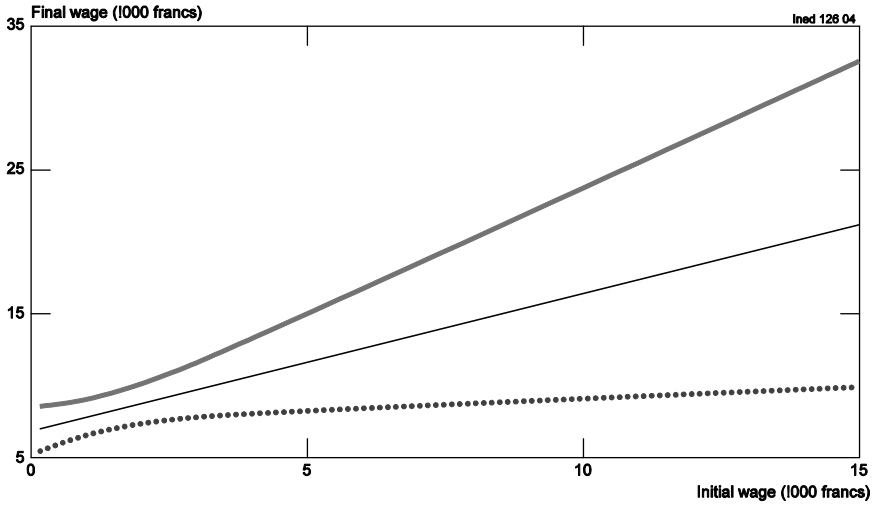


Figure VI.2. - Final wage as a function of initial wage, with a 95% confidence interval

To illustrate more clearly the area's role on the curve showing the effect of the initial wage on the final wage, we can estimate level-2 residuals for each area (see equation VI.15). With these residuals, we can estimate the value of the final wage as a function of the initial wage for each département. Figure VI.3 shows these département estimates, which supply very useful information on wages of individuals present at age 15 in the départements. We have identified a certain number of départements on these regression lines that give us a better view of the differences between them. The Paris-region départements, for example, have a large y-intercept, but relatively weak slopes. By contrast, the départements of Normandy, for example, have a low y-intercept and a steep slope.

Let us now turn to the effect of the aggregate characteristic (mean initial wage in the département), introduced by model 4. Again, we find a significant and still positive effect: the mean wage will influence the final individual wage, irrespective of the initial individual wage. We have therefore identified a contextual effect: the higher the mean wage, the higher the final individual wage. The introduction of the mean wage will also reduce the regional random variable linked to the constant term, i.e., the nearly 20% mean wage increase during the observation period. By contrast, it has no impact on the regional random variable representing the initial wage or on the random variable at the individual level.

Let us now try to introduce the other individual characteristics to see their effect on the final wage. Table VI.2 reports the results.

TABLE VI.2. - EFFECT OF SELECTED CHARACTERISTICS ON THE FINAL WAGE

Parameters	Multilevel model		
Fixed	Model 5	Model 6	Model 7
a_0 (constant)	7.144 (0.444)	11.558 (0.395)	12.183 (0.400)
a_1 (initial wage)	0.889 (0.051)	0.291 (0.050)	0.475 (0.058)
a_2 (mean regional initial wage)	0.597 (0.193)	0.414 (0.155)	0.412 (0.154)
a_3 (sex)	-2.709 (0.124)	-3.570 (0.127)	-3.530 (0.128)
a_4 (unskilled manual worker)	-	-4.344 (0.199)	-4.035 (0.199)
a_5 (skilled manual worker)	-	-3.610 (0.207)	-3.302 (0.208)
a_6 (white-collar worker)	-	-2.117 (0.184)	-1.869 (0.184)
a_7 (manager)	-	3.931 (0.291)	3.578 (0.291)
a_8 (generation)	-	-	-0.223(0.022)
Random, level 2			
σ^2_{uo} (constant)	0.330 (0.129)	0.283 (0.088)	0.321 (0.101)
σ^2_{u1} (wage)	0.106 (0.027)	0.079 (0.018)	0.125 (0.028)
σ_{u13} (wage, sex)	-0.119 (0.040)	-0.109 (0.025)	-0.092 (0.028)
σ_{u15} (wage, skilled manual worker)	-	-0.165 (0.034)	-0.233 (0.054)
σ_{u18} (wage, generation)	-	-	-0.013 (0.006)
σ_{u58} (skilled manual worker, generation)	-	-	0.036 (0.018)
Random, level 1			
σ^2_{e0}	18.46 (0.383)	15.50 (0.322)	15.16 (0.315)
-2 (log-likelihood)	27605.1	26750.2	26639.8

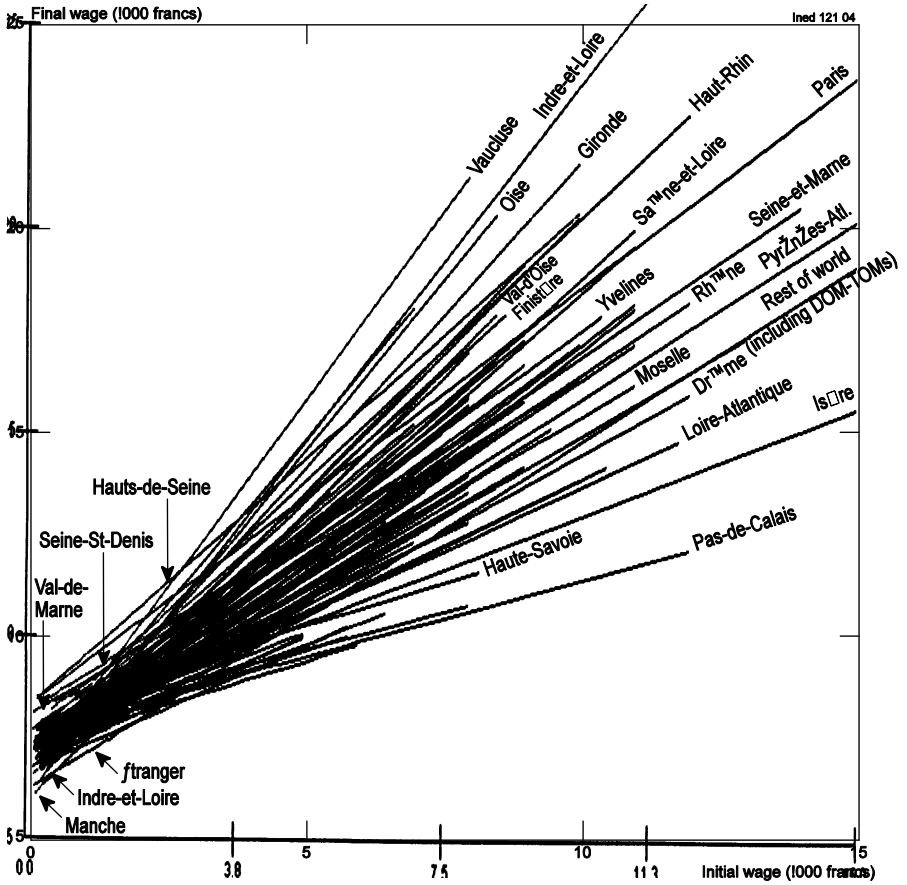


Figure VI.3. - Final wage as a function of initial wage (’000 francs)

Model 5 introduces a binary variable representing the individual’s sex. It shows that the mean income of women ($x_{3ij} = 1$) is more than 2,700 francs below that of men. As we shall see, this result is always obtained when we examine the effect of many other variables simultaneously. We can introduce such a characteristic into the random variables at the département level, with some precautions, as it is a binary variable. We can confirm the non-significance of the random variables representing the covariance between the constant term and the wage or sex. Table VI.2 accordingly includes only the covariance between the initial wage and sex, which alone is significant. This negative covariance shows that the initial wage has different effects on men and women. For men, $x_{3ij} = 0$, we can write:

$$\mathrm{var}(u_0 + u_1 x_{1ij}) = 0.330 + 0.106 x_{1ij}^2$$

(VI.17)

and for women:

$$\mathrm{var}(u_0 + u_1 x_{1ij} + u_3 x_{3ij}) = 0.330 - 0.119 x_{1ij} + 0.106 x_{1ij}^2 \qquad \text{(VI.18)}$$

While the first curve is uniformly increasing, the second dips to around 560 francs. The introduction of this characteristic also reduces the individual random variable by about 10%.

Model 6 introduces the individual's socio-occupational category at the time of his or her first permanent job. It is now a polytomous characteristic that we have defined with five groups of occupations: unskilled manual workers, skilled manual workers, white-collar workers, intermediate occupations, and managers. To introduce such a characteristic, we need to designate one of the groups as the control group: we have chosen intermediate occupations. Again, we observe a very significant effect of occupation, which introduces a continuous increase in the final wage from unskilled manual workers to managers. Its introduction sharply reduces the effect of the initial wage, to which it is obviously linked. However, the initial occupation explains the final wage better than the initial wage does. The initial occupation will also affect random variables at both the département level (halving the variance of the effect of the initial wage) and the individual level (15% reduction). At the département level, we find an interaction between the fact of being a skilled manual worker and the initial wage, of the same type as the interaction between sex and wage examined earlier: for women, for example, the variance at the département level is minimal for an initial wage of about 1,750 francs.

Model 7 introduces the generation. As already noted, we are working here on 10 generations, which means that younger workers should have smaller wage increases than older ones. This is indeed what we observe in the final model. The wages of the oldest individuals are 2,230 francs higher than those of the younger ones. The model entails only a mild decrease in the individual random variable, but shows new effects on the département random variable. For example, for men of the 1962 generation who are not skilled manual workers, the variance at the département level is written:

$$\text{var}(u_0 + u_1 x_{1ij} + u_8 x_{8ij}) = 0.321 + 0.125 x_{1ij}^2 - 2 * 0.013 * 10 * x_{1ij} \qquad \text{(VI.19)}$$

and is minimal for an initial wage of about 900 francs.

III. Risks of erroneous inference

We shall now examine some problems that may arise when estimating multilevel models. The first is the imputation to a given aggregation level of the effect of individual characteristics not introduced into the model, even as that aggregation level plays no role in the phenomenon studied. The second problem is the incorrect imputation to a given aggregation level of characteristics operating at another aggregation level.

Imputation to a given level of the effect of omitted fixed characteristics

Let us suppose that we have estimated a classic linear-regression model with one characteristic studied and a single explanatory variable. We know that if a second explanatory variable is independent of the first, then, in a new regression including both characteristics, the parameter estimated for the first will remain unchanged.

In a multilevel linear regression, the problem will grow more complicated, because we are now dealing with several aggregation levels and a given characteristic can be independent of another at one of these levels but be dependent on it at another level. Let us see more specifically what happens when there are two aggregation levels.

Let us suppose that the second characteristic, x_{2ij}, is independent of the first, x_{1ij}, at the aggregated level but dependent on it at the individual level. We shall see with a practical example how this can happen. In this case the model to be estimated should include random

variables only at the individual level, but none at the aggregated level, and should be written as follows:

$$y_{ij} = a_0 + e_{0ij} + (a_1 + e_{1ij})x_{1ij} + (a_2 + e_{2ij})x_{2ij} + (a_{12} + e_{12ij})x_{1ij}x_{2ij} \qquad \text{(VI.20)}$$

Where e_{0ij}, e_{1ij}, e_{2ij}, and e_{12ij} are zero-mean random variables. Their variance-covariance matrix cannot be estimated with a classic regression model. As shown later, this is possible with a multilevel model, where the level-2 random variables are all null.

If we have no measure for the variable x_{2ij}, then we can only estimate a model in which the sole variable is x_{1ij}. We can easily establish that the following relationships obtain:

$$a_0 + e_{0ij} \le a_0^{'} + e_{0j}^{'} \le a_0 + a_2 + e_{0ij} + e_{2ij}$$

$$a_1 + e_{1ij} \le a_1^{'} + e_{1j}^{'} \le a_1 + a_{12} + e_{1ij} + e_{12ij}$$

where $e_{0j}^{'}$ and $e_{1j}^{'}$ are random terms that will therefore now depend on the level-2 unit. We can thus rewrite the previous model as:

$$y_{ij} = a_0^{'} + e_{0j}^{'} + (a_1^{'} + e_{1j}^{'})x_{1ij} + e_{ij} \qquad \text{(VI.21)}$$

The multilevel estimation of this model will therefore provide level-2 random variables, whereas these should be null. The outcome will be an erroneous inference. The following simulated example will provide a clearer picture of the situation.

Example no. 7: Comparison of pupils' scores at two dates by means of a simulation[10]

As in the very didactic example given by Woodhouse et al. (1996), we suppose that we measure progress in a given subject by means of tests taken at ages 8 and 11, which we examine at the individual level and the school level. Here, we assume that progress does not depend on the school but only on a second characteristic of pupils: parental support (coaching) for the subject. Parental support is measured by a variable binary, equal to unity if support is provided and zero if not. The support is all the more significant if the child has obtained a low score at age 8 but, again, this interaction is independent of the school attended.

In these conditions, we can simulate samples in which all these hypotheses obtain, with random parameters that are independent of the schools and lie in the following intervals:

$$2 \le a_0 + e_{0ij} \le 7$$

$$0.25 \le a_1 + e_{1ij} \le 1.25$$

$$21 \le a_2 + e_{2ij} \le 29 \qquad \text{(VI.22)}$$

$$-0.57 \le a_{12} + e_{12ij} \le -0.43$$

The last relationship shows us that the interaction between parental support and test score at age 8 is negative and equal, on average, to –0.5.

[10] This comparison was made by Courgeau and Baccaïni (1997).

We report below the results obtained on one of these samples, which closely resemble the results of the other samples. Let us begin by assuming that we confine our measurement at the school level to scores at ages 8 and 11, parental support being unknown. Table VI.3 gives the result of the analysis of these data.

TABLE VI.3. - PARAMETERS ESTIMATED IN MULTILEVEL MODEL LINKING SCORE AT AGE 11 TO SCORE AT AGE 8

Parameters	Estimations (standard deviation in parentheses)
Fixed	
a_0' (constant)	16.720 (1.189)
a_1' (score at age 8)	0.503 (0.033)
Random variables, school level	
σ_{u0}^2	57.000 (14.080)
σ_{u01}	-1.298 (0.373)
σ_{u1}^2	0.030 (0.011)
Random variables, pupil level	
σ_{e0}^2	91.730 (2.977)
-2 (log-likelihood)	14781.1

This table shows a very significant school effect, which is very close to the one obtained by Woodhouse et al. (1996). Figure VI.4, which plots level-2 variance as a function of the initial score, shows that the higher the child's score, the less the score at age 11 will depend on the grade obtained at age 8: the values range from 45 for an initial score of 5 points to almost 1 for an initial score of 40. To visualize the relative position of schools more precisely, we plotted on figure VI.5 the score at age 11 as a function of the score at age 8 for each school observed. Some schools appear to obtain a good performance from all their pupils in the subject tested, regardless of their initial score, whereas others seem to neglect pupils with a weak initial level.

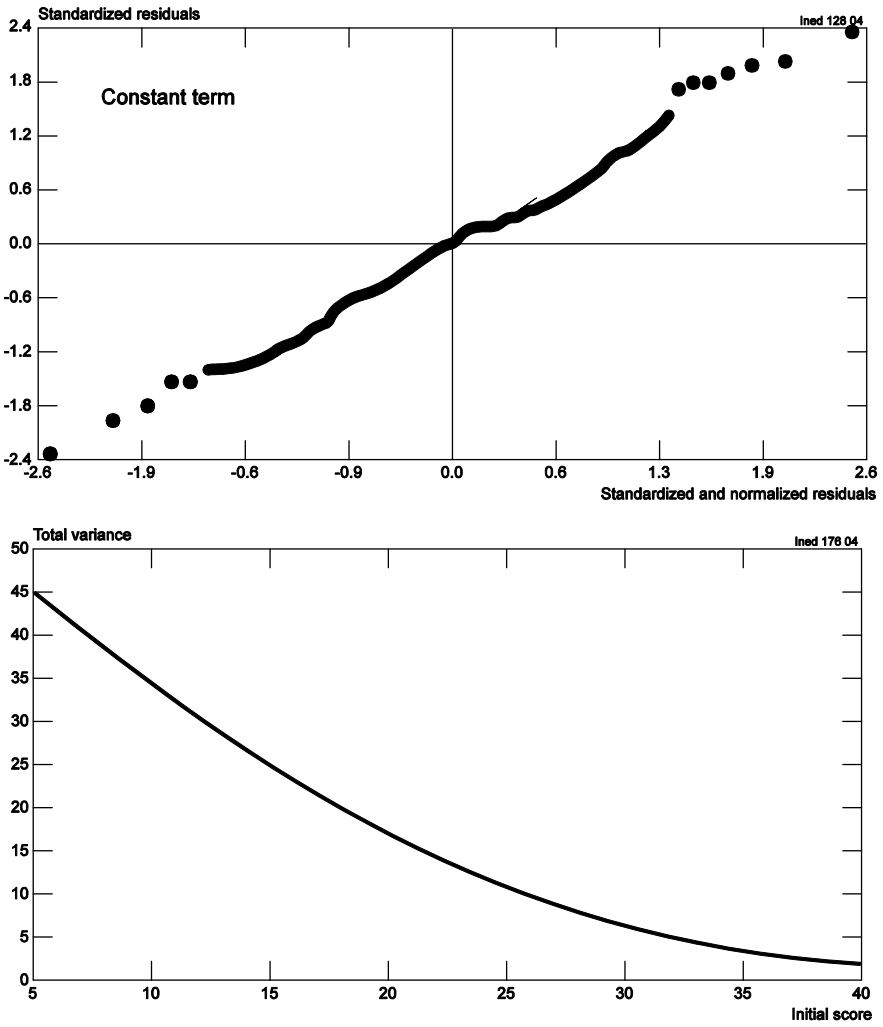


Figure VI.4. - Variance estimated at school level as a function of score

Let us now suppose that a parent survey enables us to determine which children are supported by their parents in this subject. We incorporate into the multilevel model both the parental support measured by a dichotomous variable and the interaction between support and score at age 8. The model's new parameters are given in table VI.4.

TABLE VI.4. - PARAMETERS ESTIMATED IN THE MULTILEVEL MODEL LINKING SCORE AT AGE 11 TO SCORE AT AGE 8, WITH PARENTAL SUPPORT

Parameters	Estimations (standard deviation in parentheses)
Fixed	
a_0 (constant)	4.410 (0.547)
a_1 (score at age 8)	0.766 (0.023)
a_2 (parental support)	25.170 (0.783)
a_3 (interaction between score at age 8 and parental support)	-0.529 (0.033)
Random variables, school level	
σ^2_{u0}	0.000
σ_{u01}	0.000
σ^2_{u1}	0.000
Random variables, pupil level	
σ^2_{e0}	54.840 (1.757)
- 2 (log-likelihood)	13650.3

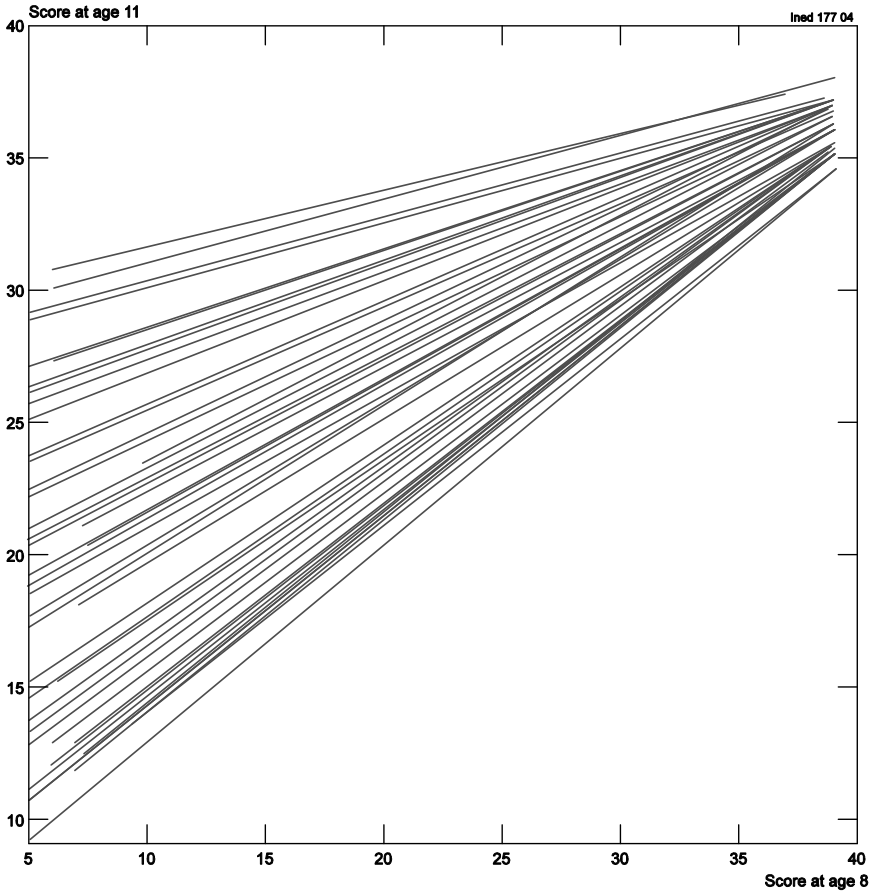


Figure VI.5. - Estimated relationships between scores at ages 8 and 11 in each school, for a multilevel model applied to a simulated sample of schools

We see that all the random variables at the school level cancel out and that the random variable at the individual level is nearly halved. All the fixed effects are significant and we find the same value of –0.5 for the interaction between the age-8 score and parental support, disregarding the random variables.

In sum, the risks of erroneous inference can be substantial when using a multilevel model if we impute to the wrong level—here, the school level—effects that actually operate at the individual level, on the model's fixed parameters. A sensible precaution is to incorporate into the fixed part the largest possible number of characteristics affecting the phenomenon. This will minimize the risk of attributing to an aggregation level an effect that does not exist in reality.

Imputation of random effects to an incorrect aggregation level

We noted earlier some hypotheses of a multilevel linear-regression model: the residuals must obey a Normal distribution with constant variances, which must, in particular, be independent of the model's explanatory characteristics. The means to test this is to plot on a graph the residuals of the different aggregation levels as a function of the values that would be

given by a normal distribution. When we obtain an approximately straight line, we can regard the residuals as more or less Normally distributed. Otherwise, we need to waive the constant-variance hypothesis.

One remedy for this drawback is to estimate random variables that are more complex at the individual level. The random variable ε_{0ij} may itself depend on certain individual characteristics z_{lij}, some of which may be identical to the characteristics x_{lij} introduced earlier in model VI.7. The new model can be written:

$$y_{ij} = a_0 + \sum_{k=1}^{m} a_k x_{kij} + (u_{0j} + \sum_{k=1}^{m_1} u_{kj} x_{kij} + \varepsilon_{0ij} + \sum_{l=1}^{n} \varepsilon_{lij} z_{lij}) \qquad \text{(VI.23)}$$

In addition to the previous parameters, there are now $\frac{n(n+1)}{2}$ variances and covariances to be estimated for level 1. Again, some of these variances and covariances may be null or not significantly different from zero, which means that *n* will be smaller than or equal to *m*.

When estimating a model with random variables at the individual level, the random variables estimated previously at higher aggregation levels may decrease and even disappear altogether. If so, we can conclude that these random variables were imputed to an incorrect aggregation level, for they are explained entirely by random variables at the individual level.

The best way to see this is to return to example no. 6, where we introduced characteristics only at the aggregated level of the départements.

Example no. 6 (continued): Wages in survey as a function of initial wage, in France

Let us resume our examination of model 3, where the explanatory characteristics were the constant term and the initial wage, both characteristics also operating at level 2. We can check whether the random variables obtained at the regional level and at the individual level effectively display an approximately Normal distribution. To do this, we have plotted these standardized residuals classified in increasing order as a function of what a Normal distribution would produce. This type of graph is called a normal quantile-quantile plot (QQ plot) of residuals.

Figure VI.6 shows the results for the département level. On the whole, the residuals are Normally distributed, for both the constant term and the initial wage. Figure VI.7 displays the corresponding results at the individual level: here, the curve shows that it is impossible to regard the residuals as Normally distributed. Consequently, it is useful to see if the initial wage can explain this non-Normal distribution.

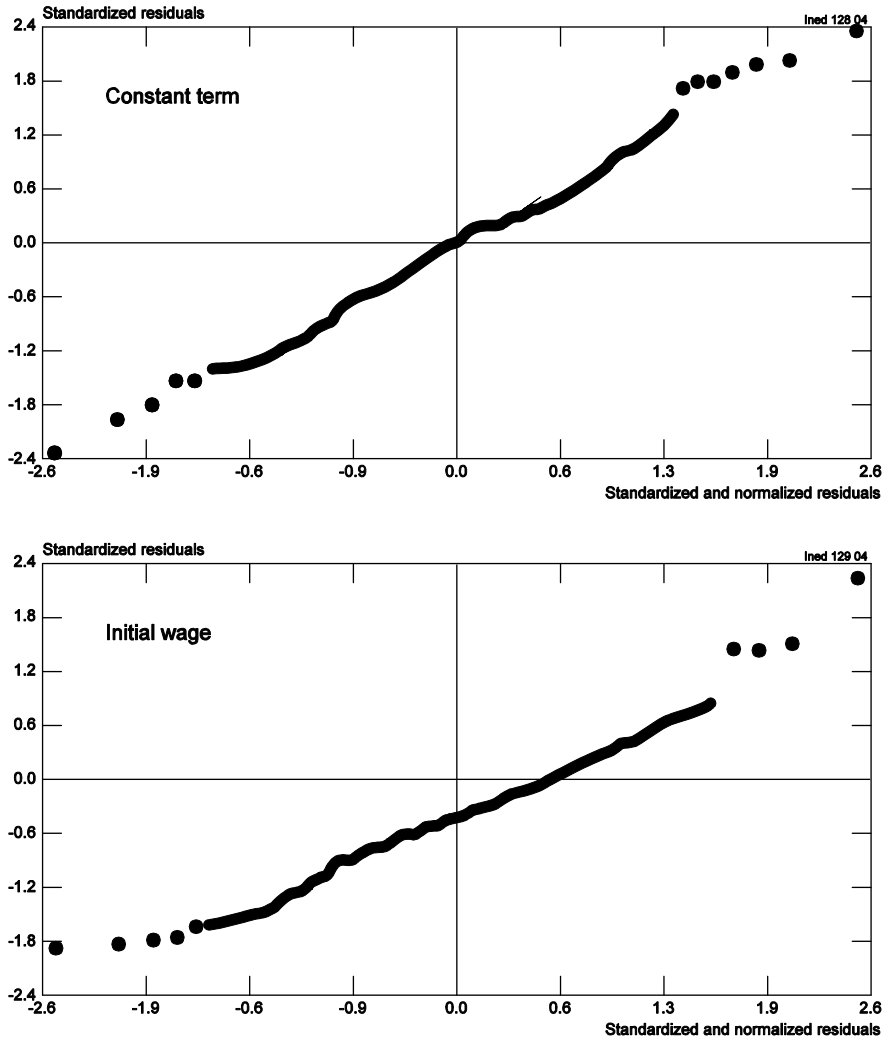


Figure VI.6. - Check for normality of residuals at *département* level

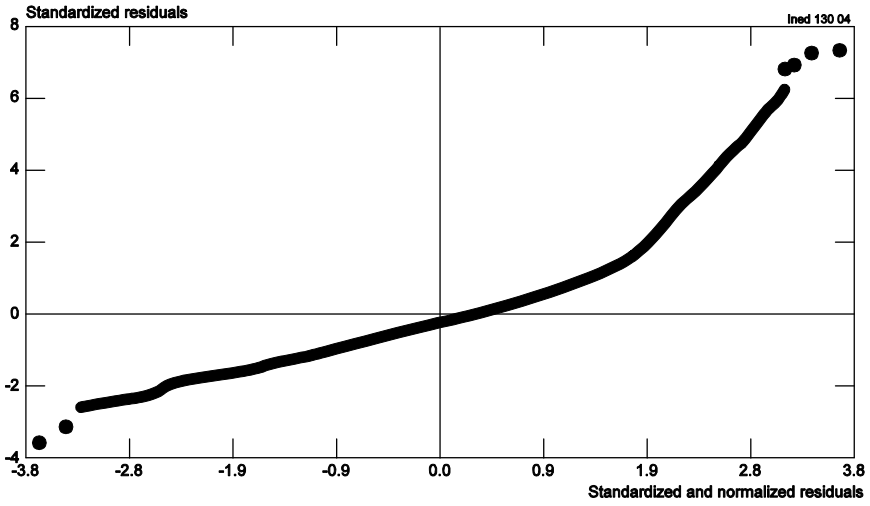


Figure VI.7. - Check for normality of residuals at individual level for constant term

The multilevel model allows us to introduce these individual random variables. The results are listed in table VI.5, which extends the analysis reported in part II of this chapter.

TABLE VI.5. - EFFECT OF VARIOUS CHARACTERISTICS OPERATING AT THE INDIVIDUAL OR AGGREGATED LEVEL ON THE FINAL WAGE (STANDARD DEVIATION IN PARENTHESES)

Parameters	***Multilevel models***			
Fixed:	Model 8	Model 9	Model 10	Model 11
a_0 (constant)	7.178 (0.126)	7.198 (0.423)	11.553 (0.376)	11.700 (0.385)
a_1 (initial wage)	0.859 (0.045)	0.776 (0.043)	0.201 (0.041)	0.415 (0.045)
a_2 (average regional initial wage)	-	0.660 (0.186)	0.536 (0.140)	0.533 (0.141)
a_3 (sex)	-	-2.664 (0.114)	-3.365 (0.113)	-3.278 (0.112)
a_4 (unskilled manual worker)	-	-	-4.497 (0.194)	-3.953 (0.202)
a_5 (skilled worker)	-	-	-3.824 (0.209)	-3.305 (0.214)
a_6 (white-collar worker)	-	-	-2.447 (0.195)	-2.039 (0.202)
a_7 (manager)	-	-	3.079 (0.360)	3.051 (0.374)
a_8 (generation)	-	-	-	-0.184 (0.018)
Random, level 2:				
σ^2_{u0} (constant)	0.491 (0.129)	0.393 (0.165)	0.312 (0.098)	0.362 (0.101)
σ_{u02} (sex, constant)	-	0	-0.112 (0.056)	-0.137 (0.055)
Random, level 1:				
σ^2_{e0} (constant)	13.155 (0.839)	11.005 (0.722)	21.715 (1.885)	23.270 (1.714)
σ_{e01} (constant, wage)	0.216 (0.402)	1.238 (0.375)	0.856 (0.399)	1.780 (0.385)
σ^2_{e1} (wage)	0.745 (0.144)	0.580 (0.126)	0.384 (0.099)	0.248 (0.097)
σ_{e03} (constant, sex)	-	0	-5.005 (0.883)	0
σ_{e13} (wage, sex)	-	-1.420 (0.179)	-1.010 (0.228)	-1.199 (0.161)
σ_{e04} (constant., unskilled manual worker	-	-	-8.073 (0.893)	-8.612 (0.880)
σ_{e14} (wage, unskilled manual worker)	-	-	-0.706 (0.269)	-0.949 (0.290)
σ_{e34} (sex, unskilled manual worker)	-	-	5.824 (0.809)	1.235 (0.340)
σ_{e05} (constant, skilled manual worker)	-	-	-5.733 (0.954)	-7.311 (0.965)
σ_{e15} (wage, skilled manual worker)	-	-	-1.222 (0.297)	-1.645 (0.311)
σ_{e35} (sex, skilled manual worker)	-	-	4.315 (0.886)	0
σ_{e36} (sex, white-collar worker)	-	-	-	-5.148 (0.833)
σ_{e08} (constant, generation)	-	-	-	-0.913 (0.135)
σ_{e48} (unskilled manual worker, generation)	-	-	-	0.718 (0.144)
σ_{e58} (skilled manual worker, generation)	-	-	-	0.916 (0.160)
σ_{e68} (white-collar worker, generation)	-	-	-	0.503 (0.130)

-2 (log-likelihood)	27584.0	27008.9	25834.0	25658.2

Model 8 clearly shows that the initial wage has a far stronger effect at the individual level than at the département level. Its variance and covariance with the constant term, at the département level, become not significantly different from zero (the estimated values are –0.021 with a standard deviation of 0.063 for σ_{u01} and 0.030 with a standard deviation of 0.029 for σ^2_{u1}), and we have accordingly posited them as null. This elimination of the initial-wage effect at the regional level is due to the introduction of the characteristic at the individual level. The only effect remaining at the département level is the constant-term effect, signaling that the curves plotted in figure VI.3 become parallel lines (figure VI.8).

The variance at the individual level will now depend strongly on the initial wage. Although this specification of the variance at the individual level as a second-degree function of the initial wage is identical to that of the département level in model 3, we can no longer interpret it in terms of random straight lines as before. In other words, the parameters σ^2_{e0}, σ_{e01}, and σ^2_{e1} no longer consist of variances and a covariance between parameters defining the y-intercept and the slope of a straight line. They have become coefficients of a quadratic function describing the variance at the individual level as a function of two characteristics: the constant term or the initial wage. This quadratic function–is uniformly increasing when the initial wage increases. We are therefore no longer dealing with a classic linear-regression model, in which the variance is independent of the explanatory variable.

Figure VI.9 plots the variations of the variance at the individual level as a function of the initial wage. It shows the size of the variance at the individual level when the initial wage is high: it rises to nearly 180 for an initial wage of 15,000 francs, compared with a value of under 0.5 for the variance at the département level, or 13.2 for a very low individual wage (120 francs).

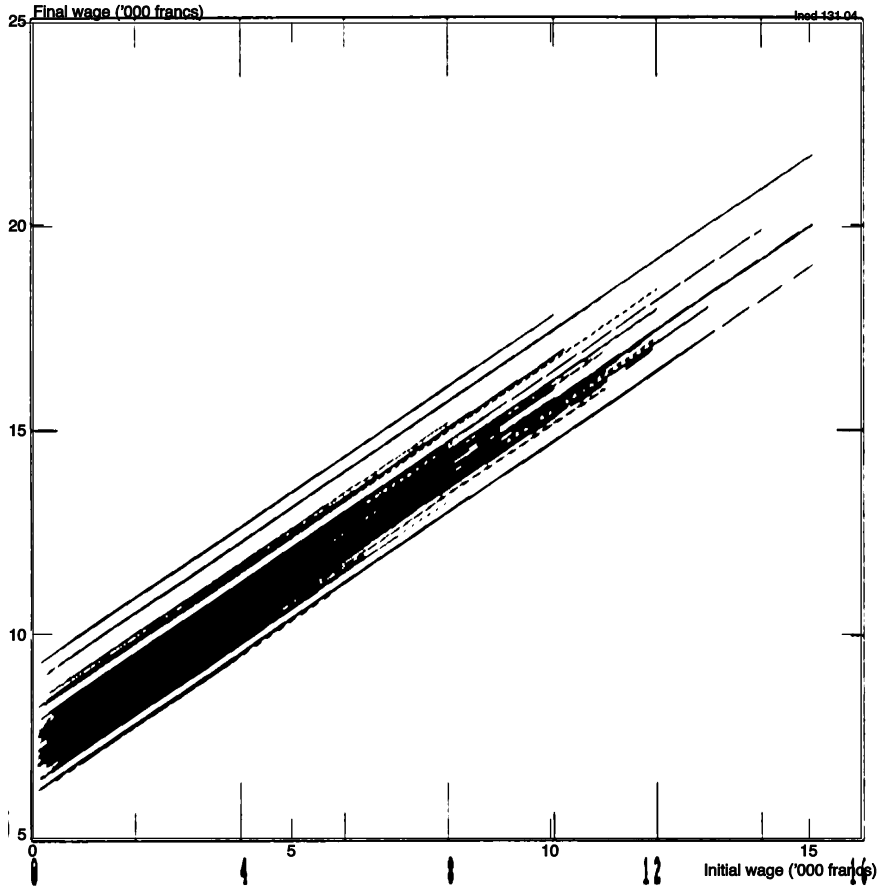


Figure VI.8. - Final wage as a function of initial wage, by area (model 8)

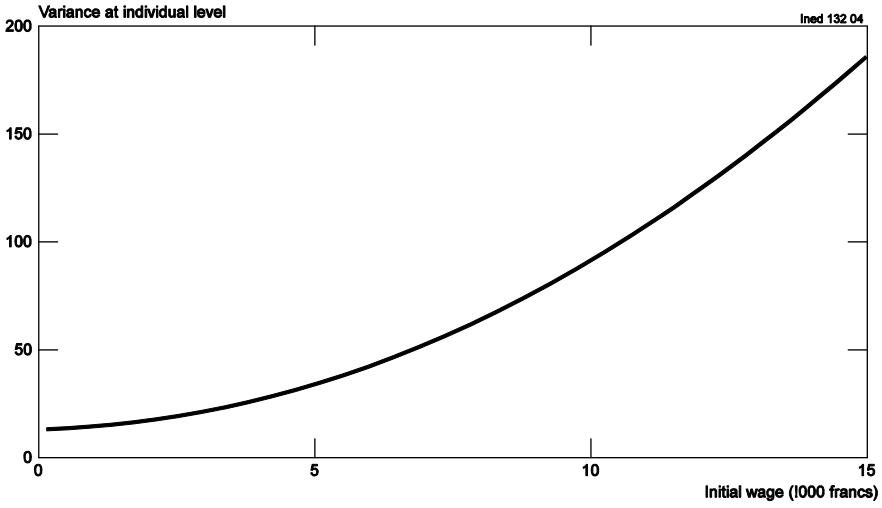


Figure VI.9. - Variance at individual level as a function of initial wage (model 8)

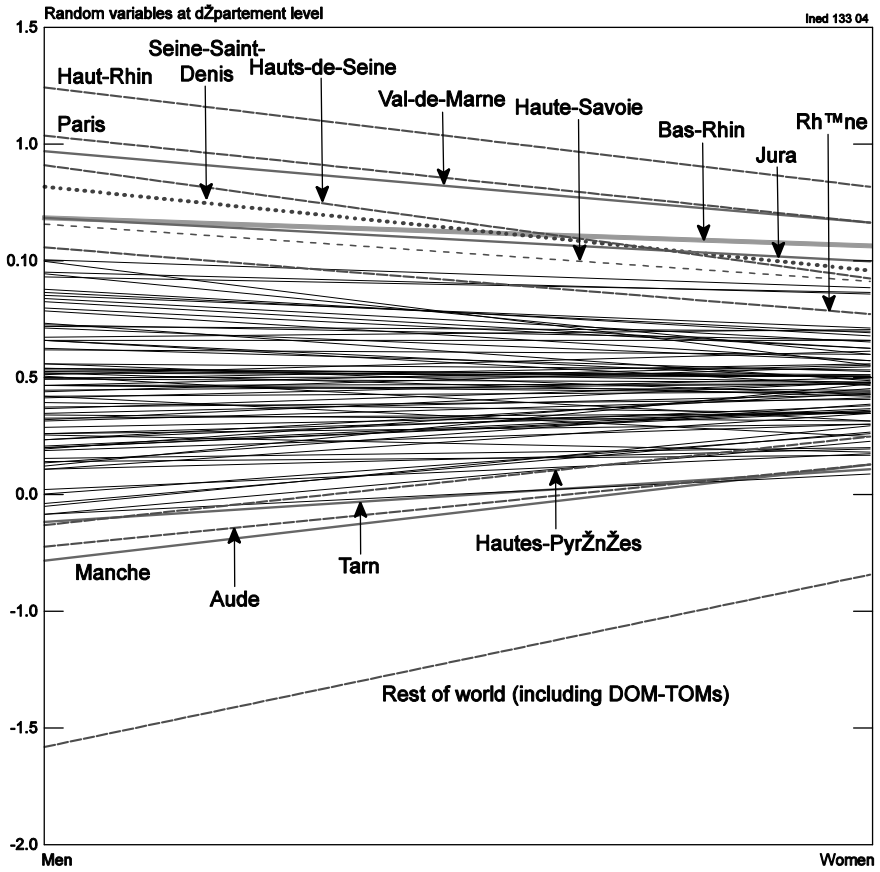


Figure VI.10. - Random variables at département level for men and women (model 8)

For the following models, we shall restrict our analysis to characteristics or variances and covariances that are significant at the 10% limit. When they are not, we posit a null effect.

Model 9—the equivalent of model 5 above, with an effect of the mean regional initial wage and a sex effect—shows little variation in the fixed parameters estimated for the initial wage and the constant. The effect of the new characteristics introduced reduces the département random variable by 15%. In contrast, variances at the individual level are fully significant, also indicating an effect of the interaction between initial wage and sex. Thus, for an initial wage of 120 francs, the individual variance of the final wage will be roughly identical for the two sexes at 11.25 for men versus 10.9 for women. For an initial wage of 10,000 francs, the individual variance of the final wage will be 93.8 for men versus 65.4 for women.

Model 10—the equivalent of model 6 above—exhibits a covariance between the constant term and sex, at the département level, whereas model 6 displayed a covariance between the initial wage and sex. This seems normal, as the initial wage no longer operates at this aggregation level. The negative covariance indicates that the dispersion of residuals at the département level is weaker for women than for men. We shall provide a graphic representation of this result when all characteristics are included (model 11). The variance at the individual level is an increasingly complex function of occupational characteristics.

Lastly, model 11—the equivalent of model 7 above—introduces the generation effect. While the effect is fully significant, it causes only a slight change in the random variable at the département level. We can calculate the residuals at the département level by sex of the individuals concerned. Figure VI.10 shows the results: the département ranking is practically

unchanged for each sex, but the variance at the département level is weaker by over one-third for women relative to men. It is quite interesting to see the départements' relative positions. The highly urbanized départements of the Paris and Lyon regions as well as the Alsatian départements have a large positive residual. This shows that—assuming all characteristics examined in this model to be equal—they register the steepest income increases. By contrast, individuals located in the rest of the world or the DOM-TOMs at age 15 register the smallest income rises. In less extreme fashion, persons located at age 15 in weakly urbanized départements such as Manche, Aude, Tarn or Hautes-Pyrénées will also exhibit low income growth.

Many terms at the individual level are added to the previous ones. Two specific examples will illustrate the magnitude of these changes in variance as a function of individual characteristics. Let us consider a woman of the generation born in 1962, with a minimal initial wage (120 francs), employed as a white-collar worker at the start of her career. The women in this category will have an individual variance of:

$$23.27 + 2 \times 0.12 \times 1.78 + 0.248 \times (0.12)^2 - 2 \times 1.199 \times 0.12 - 2 \times 5.48 - 2 \times 0.913 \times 10 + 2 \times 0.503 = 4.917$$

In contrast, a man of the generation born in 1952 with an initial wage of 10,000 francs and a first job as manager will have an individual variance of:

$$23.27 + 2 \times 10 \times 1.78 + 0.248 \times 100 - 2 \times 0.913 = 81.85$$

In other words, the individual-level variance of men in the second situation is almost 17 times as high as that of women in the first.

The likelihood statistic is far better for model 11 than for model 7, at 25610.3 versus 26639.8. However, we cannot directly conclude that model 11 is better than model 7, for the number of terms included in the first (25) is greater than those included in the second (only 15). We can use several types of criteria for this comparison. Akaike's information criterion gives:

$$-2\log_e(likelihood) + 2n$$

where n is the number of parameters included in the model. The application to both models effectively shows model 11 with 25660.3 as preferable to model 7 with 26669.8. For a discussion on the use of this criterion, see Lindsey (1999). The Bayesian information criterion gives:

$$-2\log_e(likelihood) + n\log_e(N)$$

where N is the sample size. However, in a multilevel model, it is not easy to determine whether the sample size is the number of individuals or the number of départements observed. The number of individuals is often used as a proxy for size. Here, its value is 4,777. Its application to both models again shows model 11 with 25822.1 as preferable to model 7 with 26893.9. For a discussion on the use of this criterion, see Raftery (1995).

In sum, model 11 is definitely more satisfactory than model 7. It shows the risk of inferring a fully significant effect of characteristics operating at the département level without incorporating the same characteristics simultaneously at the individual level: in the latter case the effect of certain characteristics may be nullified at the département level.

Modifying the dependent characteristic to obtain a Normal distribution of residuals

In the previous section, we showed how modeling variance at the individual level as a function of various characteristics can solve the problem of the non-Normality of residuals at the individual level. Another method consists in transforming the dependent variable and some explanatory variables so as to confirm the hypothesis of normally distributed residuals with a variance independent of the explanatory characteristics.

For non-multilevel data, various log, exponential, and other transforms have been discussed by many authors, in particular Box and Cox (1964). These procedures are also applicable to multilevel data, and we shall examine the case of a log transform on wage data.

When the measurement scale used for the dependent variable is arbitrary—as is the case with exam grades—and only the rank of individuals counts, we can try to normalize the scale in the hope that the transformation will make the residuals Normal as well. For this purpose, we classify the values of the dependent characteristic by rank and replace each value by that of a Standardized Normal distribution, whose percentage of lower values is the same as the one observed in the population. For instance, if we have a sample of 4,777 individuals, as in the example discussed throughout this chapter, the individual displaying the highest value for his or her dependent characteristic—here, 45,860 francs—will be assigned the value of 3.7075, which is the coordinate of the point x of the Standardized Normal distribution for which:

$$P(X > x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{+\infty} e^{-\frac{s^2}{2}} ds = \frac{1}{4{,}777} = 0.00021$$

and so on for the other points of the distribution. Some other explanatory characteristics, also continuous, may be replaced by equivalent variables. After these transformations, we must check that the resulting residuals are distributed normally as well.

Example no. 6 (continued): Wages in survey as a function of initial wage, in France

We noted earlier that economists prefer to use a log scale to work on wages, whose distribution is more lognormal than Normal. We shall therefore use that transform here, to see if it can normalize the distribution of residuals at the individual level. For this purpose, we apply the earlier model no. 3 to the log of the final wage explained by the log of the initial wage. We set aside the estimation of the model's parameters as well as of the residuals at the département level, which closely resemble the results in figure VI.4 estimated on raw data. Let us examine, instead, the residual's distribution at the individual level on a QQ plot (figure VI.11). We can easily see that this residual is far from being normally distributed, which means that the transform is of little value for normalizing residuals.

We must therefore consider another possible transformation: the normalization of the scales of the final wage, the initial wage, and the mean département wage. The results are reported in table VI.6.

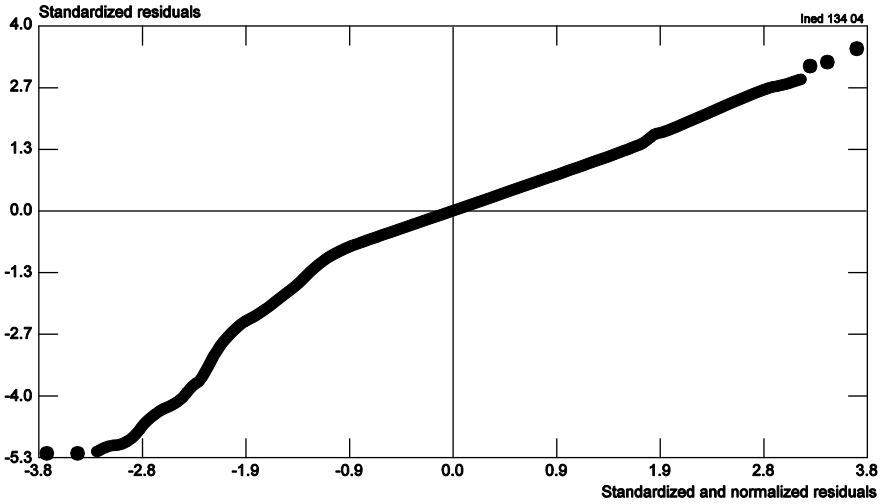


Figure VI.11. - Check for normality of residuals at individual level for the constant term (log-wage model)

TABLE VI.6. - EFFECT OF SELECTED CHARACTERISTICS OPERATING AT INDIVIDUAL OR AGGREGATE LEVEL ON NORMALIZED FINAL WAGE (STANDARD DEVIATION IN PARENTHESES)

Parameters	Multilevel models	
Fixed:	Model 12	Model 13
a_0 (constant)	-0.02105 (0.02313)	1.18500 (0.04100)
a_1 (normalized initial wage)	0.29440 (0.01382)	0.14901 (0.01568)
a_2 (normalized mean regional initial wage)	-	0.03793 (0.01816)
a_3 (sex)	-	-0.81138 (0.02637)
a_4 (unskilled worker)	-	-0.96141 (0.04051)
a_5 (skilled worker)	-	-0.76924 (0.04307)
a_5 (white-collar worker)	-	-0.49412 (0.03696)
a_7 (manager)	-	0.43915 (0.05503)
a_8 (generation)	-	-0.05054 (0.00468)
Random, level 2:		
σ^2_{u0} (constant)	0.02762 (0.00699)	0.01424 (0.003899)
σ^2_{u1} (normalized initial wage)	0.00895 (0.00394)	0
Random, level 1:		
σ^2_{e0}	0.86870 (0.01807)	0.61666 (0.01273)
-2 (log-likelihood)	13002.1	11314.0

Model 12 in this table is the equivalent of model 3 in table VI.1. The estimated parameters are naturally different, but the effect of the normalized initial wage is indeed consistent with the effect of the initial wage, and the random variables at the département level indicate the same type of parabolic variation in both models. A worthwhile exercise here is to examine the normality of the random variable at the individual level. Figure VI.12 shows the residuals' QQ plot and demonstrates their near-perfect normality.

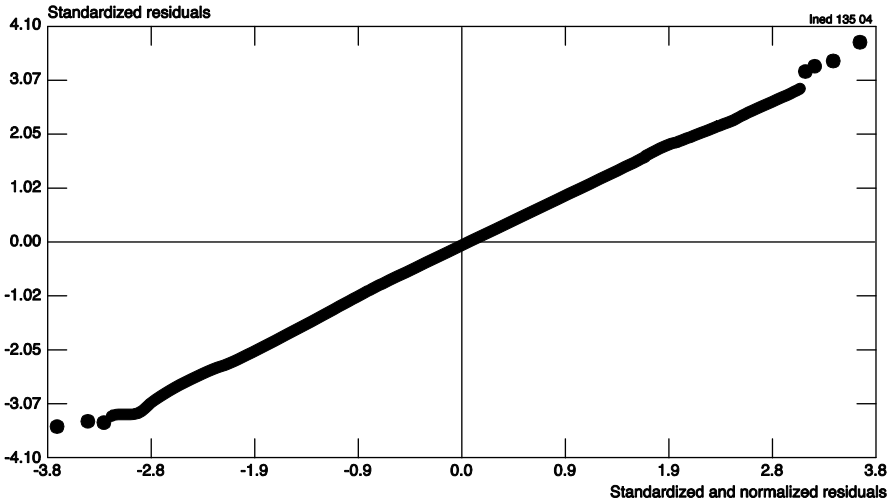


Figure VI.12. - Check for normality of residuals at individual level for the constant term (model 8)

We can then estimate the parameters of the model incorporating all characteristics, with Normally distributed residuals. Model 13 gives the results. It is important to begin by checking that the residuals are always normally distributed. We have not illustrated the curve here, as it is virtually identical to the one in figure VI.12. The test on all possible types of random variables at the département level shows that they are all non-significant except the constant term. The random variable at the département level entails a negative and significant covariance between the sex effect and the constant's effect, noted in model 11. As a result, we have omitted the random variable from model 13. The constant term's random variable is nearly halved by comparison with the previous model. By contrast, the effects of the fixed terms are consistent with those in model 11.

The parameters of model 13 are rather complex to interpret, as we are no longer modeling the final wage itself. We need to examine the curves that show wages as a function of their normalized equivalents, which we can still build, since they constitute an increasing, continuous function. But this curve no longer has a simple functional equivalent, as in the case of the previous log transform.

IV. CONCLUSION

Throughout this chapter, we have discussed a simple example of wage analysis to show the various problems posed by the multilevel analysis of continuous characteristics—and the solutions that it can provide. Such an analysis requires far greater precautions that a classic linear model.

For instance, the standard methods used to estimate a classic linear model, such as least squares, are not directly applicable to multilevel models. Successive-iteration methods allow a proxy estimation of these models' parameters under certain hypotheses, in particular that of normally distributed residuals at each level. It is important to check whether this condition is met and, if not, to attempt remedial action.

There are also many risks of erroneous inference. In a conventional regression, if we omit a characteristic that is independent of those already introduced into the model, we obtain parameters that will not change when we add the new characteristic. By contrast, in a

multilevel regression, this characteristic can be independent of the others at one aggregation level and dependent on them at another. As we have shown, this can entail the imputation of fixed effects at certain aggregation levels, whereas in fact they are non-existent there.

Likewise, the imputation to an incorrect aggregation level of characteristics operating at another aggregation level is always possible. This can happen, in particular, when the residual at the individual level is not normally distributed. If so, it is useful to introduce random variables at the individual level, as they may sometimes remove the effect of random variables incorrectly imputed to a more aggregated level.

This chapter could have addressed other problems and issues relating to the use of continuous data, such as the examination of models comprising more than two aggregation levels, the examination of individual and aggregate units supplying residuals inconsistent with the model, the various types of tests to perform on the models' hypotheses, and so on. We did not believe it would be useful to elaborate on these matters further here,[11] for, as noted earlier, such models are seldom used in demography. The events studied by demographers are typically discrete and require different estimation methods, for they need to be studied with non-linear models. That is what we shall examine in greater detail in the next chapter.

[11] For a good discussion of them, see Bryk and Raudenbush (1992).

CHAPTER VII

ANALYSIS OF DISCRETE CHARACTERISTICS

As noted in the previous chapter, the demographer works essentially on discrete data. The most satisfactory observation will be of the event-history type, which examines the occurrence of events throughout an individual's lifetime: the analytical methods used in that approach are described and discussed in chapter VIII. But often the only information available to the demographer concerns the occurrence of an event in a fixed period of time, the presence or absence of a given characteristic, or the number of events occurring in a given duration. The information may be obtained from vital statistics, examined cross-sectionally, from the census or from surveys conducted at a given moment. Such censuses or surveys may ask persons about recent events (migrants during the inter-census period, for example), their current status (family status or occupation, in the census; whether respondent is unemployed, in the labor-force survey; contraceptive method currently used, etc.) or the number of events experienced in their lifetimes (number of dwelling changes since age 15, total number of children, number of unions, etc.).

In all these cases, the dependent variable will be (1) a dichotomous variable (migrant or not, unemployed or not, etc.), (2) a polytomous variable, if several situations are possible (never-married, living in a union, married, widowed or divorced, for example), or (3) an event count (number of dwelling changes, number of children, etc.). A linear model of the type presented in chapter VI will not function in these conditions. The hypothesis of continuously distributed residuals is no longer tenable, since we are dealing with discrete variables and, a fortiori, the hypothesis of a Normal distribution. We must therefore find the right way to model these distributions of discrete variables.

In this chapter, we distinguish three major types of models corresponding to the different types of qualitative characteristics listed above: (1) model with dichotomous variables (logit model, probit model or complementary log-log models), (2) models with polytomous variables (ordered or not ordered polytomous logit model), and (3) models for event counts (log-linear model or Poisson model).

These models actually generalize the linear models used in chapter VI (for more details, see McCullagh and Nelder, 1989). They are all derived from a generalized linear model, capable of estimating these various special cases. In fact, the following chapter will show that we can use the same approach to estimate multilevel event-history models.

We begin, therefore, with an overview of the generalized linear model, before discussing the different types of modeling of discrete characteristics, illustrated by examples of specific applications.

I. - A two-level generalized linear-regression model

This type of model was progressively implemented in statistics, starting with simple cases steadily generalized to allow an estimation of its parameters in a wide range of situations. We offer a brief presentation here, referring the interested reader to the articles and books describing the model (McCullagh and Nelder, 1989; Lindstrom and Bates, 1990; Goldstein, 1991; Schall, 1991; Gumpertz and Pantula, 1992; Breslow and Clayton, 1993; Wolfinger, 1993).

We take the simplified case in which we observe only two aggregation levels, as well as the effect of a single individual characteristic and a single random variable at the aggregate level. This will allow a simpler presentation than the general case. It is easy, however, to generalize the results of this chapter to more complex cases involving several explanatory characteristics and aggregation levels, with random variables that may also be nested.

This model supposes that the expected proportion of the dependent variable, $E(y_{ij})$ (where j is the aggregation level in which individual i is located), is a non-linear function f of (1) the prediction, itself linear, based on the explanatory variable x_{1ij}, $a_0 + a_1 x_{1ij}$, and (2) the random variable , u_{0j}, operating at aggregation level j. We can therefore write the following equation, equivalent to the previous one (VI.9):

$$y_{ij} = f(a_0 + a_1 x_{1ij} + u_{0j}) + \varepsilon_{ij} = f(\eta_{ij}) + \varepsilon_{ij}$$

(VII.1)

where η_{ij} is the linear function of the explanatory variable and random variable introduced here and ε_{ij} is the model's individual random variable. We assume that both random variables are normally distributed (i.e., that $u_{0j} \approx N(0, \sigma_u^2)$ and $\varepsilon_{ij} \approx N(0, \sigma_e^2)$) and independent (i.e., that $\text{cov}(u_{0j}, \varepsilon_{ij}) = 0$). The function f is, for the moment, generic. We shall see the forms it will take in different situations analyzed in greater detail in the course of this chapter. It is also worth examining the function f^{-1}, which will transform $y_{ij} - \varepsilon_{ij}$ into the linear prediction $a_0 + a_1 x_{1ij} + u_{0j}$.

We can use various methods to estimate the model's parameters and random variables. We briefly describe an algorithm that proxies the f function with a linear model, creating a situation similar to the one addressed in chapter VI (Goldstein, 1991). Other procedures have also been proposed: for more details on them, see the publications by Longford (1987) and Raudenbush and Bryk (1986).

The method used here will consist in estimating the parameters and random variables by successive approximations, using a Taylor series expansion of the f function. The estimators' accuracy will depend on the number of terms included in the expansion. We start with the terms estimated at stage n to calculate the ones operating at stage n+1.

Let us begin by assuming that we use a first-order series expansion around the values of the fixed parameters, equal to a_0^n and a_1^n , and of the random parameter, equal here to zero. We obtain the new value of y_{ij} as a function of these parameters:

$$y_{ij} = y_{ij}^n + \left[a_0^{n+1} - a_0^n + \left(a_1^{n+1} - a_1^n\right)x_{ij} + u_{0j}\right]f'(\eta_{ij}^n) + \varepsilon_{ij} \quad \text{(VII.2)}$$

where $f'(\eta_{ij}^n)$ is the derivative of the linear prediction of the f function at the point η_{ij}^n. Writing this relationship as:

$$y_{ij} - y_{ij}^n + a_0^n + a_1^n x_{ij} = [a_0^{n+1} + a_1^{n+1}x_{ij} + u_{0j}]f'(\eta_{ij}^n) + \varepsilon_{ij}$$

(VII.3)

we see that the model becomes linear as a function of the parameters to be estimated at the following iteration: the dependent characteristic is adjusted by means of an "offset" term equal to $y_{ij}^n - a_0^n - a_1^n x_{ij}$, and the parameters to be estimated are weighted by the f function's derivative, estimated at the nth iteration. This linear model can therefore be solved using the methods presented in chapter VI. We can improve the estimate in different ways.

First, it is useful to add a second-order expansion for the random term, i.e., to include the following term in the second member of the equation (VII.2):

$$\frac{1}{2}u_{0j}^2 f''(\eta_{ij}^n)$$

(VII.4)

which allows a better estimation (Goldstein and Rasbash, 1996). Equation (VII.3) can generate large biases when the number of level-1 units per level-2 unit is low and the values of the underlying random parameters are high. This method enables us to estimate what are known as marginal quasilikelihood estimators.

A second useful procedure is to start with non-zero values for the random term. In this case, the series expansion is also performed from the $\hat{u}_{0j}$ estimate at stage n, which gives:

$$f(\eta_{ij}^n) = a_0^n + a_1^n x_{ij} + \hat{u}_{0j}$$

(VII.5)

and the relationship (VII.2) can be rewritten as:

$$\begin{aligned} y_{ij} &= y_{ij}^n + [a_0^{n+1} - a_0^n + (a_1^{n+1} - a_1^n)x_{ij} + u_{0j} - \hat{u}_{0j}]f'(\eta_{ij}^n) \\ &\quad + \frac{1}{2}(u_{0j} - \hat{u}_{0j})^2 f''(\eta_{ij}^n) + \varepsilon_{ij} \end{aligned} \quad \text{(VII.6)}$$

The addition of these terms yields a better estimate of the random variable, especially when its value is high. Under certain conditions, we can still estimate the model in the same way as the linear model in chapter VI (Goldstein and Rabash, 1996). This method enables us to determine what are called quasilikelihood estimators using a second-order approximation.

With these various approximations, we can estimate the log-likelihood maximum for a non-linear model. We can use the proxy estimators to perform certain tests and construct confidence intervals, but the approximations may yield unsatisfactory results. That is why we shall not list the –2 (log-likelihood) value for the models described below. It will be better to test whether a new characteristic introduced into the model—a fixed-effect or random-effect variable—has a significant effect. There are standard χ^2 tests for doing so.

II. - Modeling binary data

Let us suppose that, for each individual i present in an area j, we measure a dependent characteristic in binary form: either the individual has this characteristic, in which case $y_{ij} = 1$, or the individual does not, in which case $y_{ij} = 0$. We can therefore write:

$$P(y_{ij} = 1) = \pi_{ij} \qquad \text{and} \qquad P(y_{ij} = 0) = 1 - \pi_{ij}$$

(VII.7)

where π_{ij} is the probability that the individual possesses the characteristic. We write that the y_{ij} variables have a binomial distribution:

$$y_{ij} \approx Bin(1, \pi_{ij})$$

(VII.8)

The conditional variance of y_{ij} will be:

$$\text{var}(y_{ij} \left| \pi_{ij} \right.) = \pi_{ij}(1 - \pi_{ij})$$

(VII.9)

In most studies, the explanatory characteristics at our disposal for each individual may be binary (the individual is a farmer or not), polytomous (the individual is never-married, living in a union, married, widowed, divorced or separated), or continuous (the individual earns a given wage). When the variable is polytomous with n categories, we often replace it by (n-1) binary variables, one for each category compared with the category that is not introduced. We can represent these p explanatory characteristics in the form of a matrix with a general term x_{pij}. The study's main purpose will be to analyze the relationship between the probability π_{ij} and the characteristics x_{pij}.

Sometimes, we can aggregate individuals possessing various values for their explanatory characteristics. We can still use the same formulations, noting that the measures are performed here on group i and not on individual i. The size of the i^{th} group in area j is m_{ij}. If the observations are independent of one another, for all individuals, and if the probability that individuals of the same group possess the dependent characteristic is constant in each group, then the distribution of the y_{ij} variables will be binomial with the following parameters:

$$y_{ij} \approx Bin(m_{ij}, \pi_{ij})$$

(VII.10)

and its conditional variance will be equal to:

$$\text{var}(y_{ij} \left| \pi_{ij} \right.) = \frac{\pi_{ij}(1 - \pi_{ij})}{m_{ij}}$$

(VII.11)

However, while observations in different groups can generally be assumed to be independent, observations within the same group can more often be correlated. We thus need to introduce an "extra binomial" variation, which multiplies the estimated conditional variance by a parameter σ^2 to be estimated.

The omission of an aggregation level in the analysis may introduce a new "extra binomial" variation. For example, the "Youth and Careers" survey, which interviewed the members of the same household aged over 16, when we did not take this selection criterion into account, this introduced a new "extra binomial" variation. In other circumstances we may sum—for a given area—separate binomial variables with different probabilities, which will generate an even more complex structure for the conditional variance.

Continuing with our simplified example, let us see how to model the influence of a fixed characteristic and a random characteristic on the behavior of individuals (or groups) situated in different areas. This gives a model with two aggregation levels. Again, we can easily generalize this situation to more complex conditions and to a larger number of aggregation levels.

To search for relationships between the dependent characteristic and explanatory characteristics, it is useful to set up a model capable of writing them. A linear model can be very useful here, but the probability π_{ij} cannot be expressed directly as a linear function of these characteristics: a linear function may produce probability estimates lying outside the limits (0,1) imposed by probability theory. One way to avoid this inconvenience is to use a f^{-1} transform, which maps the interval (0,1) onto the entire real line $(-\infty,+\infty)$. Three functions are typically used to achieve this:

$f^{-1}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$ called the logit function

$f^{-1}(\pi_{ij}) = \Phi^{-1}(\pi_{ij})$ called the probit function, where Φ^{-1} is the inverse of the integral of the standardized Normal distribution

$f^{-1}(\pi_{ij}) = \log[-\log(1-\pi_{ij})]$ called the complementary log-log function.

The differences between logit function and probit function are minimal when $0.1 \le \pi_{ij} \le 0.9$, so that we usually cannot distinguish between them in this interval. For low values of π_{ij}, the complementary log-log function cannot be distinguished from the logit function. We shall continue this presentation by working on the logit model, whose interpretation in probabilistic terms is simple, but we must keep the other models in mind when modeling a binary variable.

Here, we can write:

$$f^{-1} = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = a_0 + a_1 x_{1ij} + u_{0j}$$

(VII.12)

and we can easily check that the variable f equal to π_{ij} is written:

$$f = \pi_{ij} = \left\{1 + \exp\left(-\left[a_0 + a_1 x_{1ij} + u_{0j}\right]\right)\right\}^{-1}$$

(VII.13)

The functions f' and f'' are written:

$$f' = f\left\{1 + \exp\left(a_0 + a_1 x_{1ij} + u_{0j}\right)\right\}^{-1}$$

(VII.14)

$$f'' = f'\left\{1 - \exp\left(a_0 + a_1 x_{1ij} + u_{0j}\right)\right\}\left\{1 + \exp\left(a_0 + a_1 x_{1ij} + u_{0j}\right)\right\}^{-1} \quad \text{(VII.15)}$$

Going back to the observed binary characteristic y_{ij}, we see that we can write:

$$y_{ij} = \pi_{ij} + e_{ij}\sqrt{\frac{\pi_{ij}(1-\pi_{ij})}{m_j}} \quad \text{where } m_j = 1 \text{ in the case of individual measures} \quad \text{(VII.16)}$$

The variable e_{ij} is not normally distributed here, for when $y_{ij} = 1$ with the probability π_{ij}, the variable is equal to $e_{ij}^1 = \sqrt{\frac{1-\pi_{ij}}{\pi_{ij}}}$, and when $y_{ij} = 0$ with the probability $(1-\pi_{ij})$, it is equal to $e_{ij}^0 = -\sqrt{\frac{\pi_{ij}}{1-\pi_{ij}}}$. We confirm that:

$$E(e_{ij}) = \pi_{ij}e_{ij}^1 + (1-\pi_{ij})e_{ij}^2 = 0 \qquad \text{and} \qquad \text{var}(e_{ij}) = \pi_{ij}\left(e_{ij}^1\right)^2 + (1-\pi_{ij})\left(e_{ij}^2\right)^2 = 1$$

This model is therefore a non-linear model that we can estimate using the procedure described in the first part of the chapter. For individual measures, we can write the first-order Taylor series expansion:

$$y_{ij} = (a_0 + a_1 x_{1ij} + u_{0j})f' + e_{ij}\sqrt{\pi_{ij}(1-\pi_{ij})} \quad \text{(VII.17)}$$

so that, for a given value of x_{1ij}, we can write:

$$\text{var}(y_{ij} \mid x_{1ij}) = \sigma_{u0}^2 \pi_{ij}^2 \left\{1 + \exp\left(a_0 + a_1 x_{1ij}\right)\right\}^{-2} + \pi_{ij}(1-\pi_{ij}) \quad \text{(VII.18)}$$

For more details on this estimation, we refer the interested reader to Goldstein (1991, 2003).

To show the practical results of these estimations, we begin by returning to the example concerning Norwegian migrations, discussed through part I of this book.

Example no. 2 (continued): Migrations in Norway as a function of various characteristics

Let us take up the analysis where we left off in chapter IV. The only characteristic examined was farmer status, with the contextual effect of the percentage of farmers in each region. We can now note that these estimates were calculated with the most sophisticated model, which uses, for example, quasilikelihood estimators under a second-order approximation. As we shall now add other individual characteristics, it is better not to take non-farmers as the base category. We now introduce a constant term, both in the fixed terms and in the random variables. It will estimate the parameter for individuals possessing none of the characteristics considered. By contrast, as this term, a_0, also applies to individuals displaying a given characteristic, we shall need to add it to the term linked to this characteristic, a_1, in order to measure the characteristic's real fixed effect. Likewise, the variance of the term expressing the characteristic's random effect will have to be calculated as the sum $\sigma_{u0}^2 + 2\sigma_{u01} + \sigma_{u1}^2$. We shall give a practical example of the application of this method later on.

Let us rewrite the model presented in the last column of table IV.3 (contextual multilevel model), substituting the constant term for the non-farmer term. The rewritten model is shown in the second column of table VII.1. The result is identical for the constant term of this table and the non-farmer term of table IV.3: –2.066. With table VII.1, we can compute the fixed effect for farmers as the sum –2.066 + 0.069 = –1.997, which is very close to the estimate in table IV.3: –2.003. Likewise, the random variable for non-farmers in table IV.3 is identical to the random variable for the constant term of table VII.1. To determine the random variable for farmers in table IV.3, we need to calculate the sum:

0.053 + 2 x 0.039 + 0.082 = 0.213, which is very close to the result in table IV.3: 0.214. Lastly, the random variable for the covariance between farmers and non-farmers is given by the sum: 0.053 + 0.039 = 0.092, identical to the result in table IV.3.

The model did not include an extra binomial variance, the first-level random variable being frozen at unity. We can determine if the logit-model hypothesis is confirmed by freeing the random variable at this aggregation level. Model 2 of table VII.1 shows the parameters estimated in this manner. We see clearly that the aggregate-level random variable does not differ significantly from unity—a perfect corroboration of the logit-model hypothesis. We can therefore set this random variable to unity in our further analysis. The effect of the other characteristics is also the same as in model 1.

TABLE VII.1. - MIGRATIONS IN NORWAY AS A FUNCTION OF VARIOUS CHARACTERISTICS, WITH STANDARD DEVIATION OF THEIR EFFECTS IN PARENTHESES

Characterisation	Multilevel logit Models			
Fixed :	Model 1	Model 2	Model 3	Model 4
a_0 (constant)	-2.066	-2.066	-1.910	-1.901
	(0.111)	(0.111)	(0.106)	(0.105)
a_1 (farmer)	0.069	0.069	-0.254	-0.254
	(0.248)	(0.248)	(0.213)	(0.213)
a_2 (percentage.of farmers)	5.394	5.394	6.646	6.511
	(1.461)	(1.461)	(1.368)	(1.349)
a_3 (farmer x percentage of farmer)	-8.871	-8.872	8.573	-8.570
	(2.959)	(2.958)	(2.457)	(2.456)
a_4 (married with children)	-	-	0.149	0.903
	-	-	(0.047)	(0.354)
a_5 (cohabiting with children)	-	-	0.329	0.333
	-	-	(0.105)	(0.105)
a_6 (married without children)	-	-	0.885	0.882
	-	-	(0.055)	(0.055)
a_7 (+12 years' education)	-	-	0.677	0.677
			(0.120)	(0.119)
a_8 (economically active)			-0.584	-0.582
			(0.065)	(0.064)
a_9 (married with children x percentage of farmers)			-	-4.390
				(2.047)
Random level 2	0.053	0.053	0.040	0.039
σ^2_{u0}	(0.019)	(0.019)	(0.016)	(0.016)
σ_{u01}	0.039	0.039	0	0
	(0.028)	(0.028)	0	0
σ^2_{u1}	0.082	0.082	0.171	0.169
	(0.074)	(0.074)	(0.086)	(0.085)
σ^2_{u7}	-	-	-0.082	-0.081
	-	-	(0.037)	(0.036)
σ_{u78}	-	-	0.052	0.050
			(0.024)	(0.023)
σ^2_{u8}				
Random level 1	1.000	0.999	1.000	1.000
	(0.000)	(0.008)	(0.000)	(0.000)

Model 3 of the same table now incorporates the other characteristics, without other contextual terms, to be examined in model 4. We have now eliminated the farmer random variable, which already seemed to have little significance in the previous models and now becomes totally insignificant. By contrast, we have introduced new random variables that seem significant. They have been noted with the number(s) of the corresponding fixed variable(s). For example, the random variable σ^2_{u8} corresponds to variable no. 8: "economically active." We have also introduced the characteristics "married with child(ren)," "married without children," and "cohabiting with child(ren)," for the original "married" and "with child(ren)"

variables allowed different behaviors of various earlier groups to escape analysis. All these effects are significant: for instance, the economically active have a much lower mobility than the other categories, demonstrating the role of unemployment or economic inactivity at age 21 on migration. Few significant new random variables appear at the regional level. The only ones to introduce differences are "over 12 years' education" and "economically active." If we could determine the regional residuals for these characteristics, we could rank the regions by their residuals. The value of the random variable for the constant term is 25% lower than in the previous model.

Model 4 adds the contextual terms for the characteristics other than "farmer." None of the characteristics included in the level-2 random variables has contextual terms allowing the variance to be reduced as in the case of farmers. Only "married with child(ren)" has a contextual effect: when the proportion of married persons with child(ren) increases in the region of residence, then the slightly higher probability of migrating for married persons with child(ren) compared with non-married persons without children will decrease. But it is important to note that the number of married persons with child(ren) is still low at the ages examined here: they represent between 13% and 21% of the regional population.

To allow a comparison with the polytomous models in the following part, we now present a second application of the logit multilevel model to the data from the "Young People and Careers" survey. The Norwegian data do not allow a finer geographic segmentation than the division into regions. For reasons of statistical confidentiality, the file that the Norwegian statistical offices allowed us to extract from the population register does not indicate parishes of residence, for example. But the use of a polytomous model requires information on the number of migrants across the different levels of territorial units. The "Young People and Careers" survey does not record the exact localities where respondents have lived, but it does indicate, when a migration has taken place, the type of geographic unit that has been crossed: municipality, département, region or country. In this first example, we compare individuals who have remained sedentary or have migrated within their municipality to other categories, and we go on to examine whether the behavior of the sedentary is indeed very similar to that of intra-municipal migrants.

Example no. 8: Migrants in year before "Young People and Careers" survey in France

We use the data from the "Young People and Careers" survey conducted in France in 1997. It asked a question on the place of residence one year earlier, with a measure of occupation at that date and a measure of selected family characteristics throughout the respondents' lives: dates of entry into union, marriage, separation, and birth of children. We shall examine these events here when they occur in a two-year period before migration. Individuals are observed from their 21st birthday to the survey date: they therefore belong to the generations born between 1952 and 1975. The geographic segmentation is identical to the one used in example no. 5 (chapter VI) on wages.

As before, we examine models incorporating a growing number of characteristics to study the behavior of individuals. We model the behavior of individuals who have not migrated outside their municipality of residence. We regard them here as sedentary, although we shall check later on whether intra-municipal migrants do not already behave very differently from non-migrants. This was the underlying hypothesis in the French 1962 census, where the only individuals classified as migrants were those who had at least moved to another municipality. We test whether the differences between French départements remain significant when we

introduce the various characteristics, without attempting to add other variables into the random terms.

Table VII.2 reports the models' results. Model 1 gives a 0.905 probability of sedentarity for the total population. The range of values across départements is rather wide, from a low sedentarity of 0.827 in Paris to a high sedentarity of 0.941 in Pas-de-Calais, creating a very significant variance at the département level. The variance of the random variable at the individual level is set at unity for a logit model. The second model frees us from this constraint and allows us to test the model's hypotheses. The variance of the random variable is categorically not significantly different from unity. This confirms the validity of a logit model, which we shall therefore use in the rest of this analysis. The parameters estimated are, in fact, identical to those given by model 1.

Model 3 introduced the individual's sex and occupation at the start of the observation period. We have classified the population here into a group of economically inactive persons and seven broad socio-occupational categories: farmers, artisans, higher-level managers, intermediate occupations, white-collar workers, skilled manual workers, and unskilled manual workers. As noted earlier, we must choose one of these categories as the base category for ranking the others: we have selected intermediate occupations. Managers and the economically inactive are the least sedentary by comparison with farmers and, to a lesser extent, artisans. White-collar workers are indistinguishable from the benchmark category of intermediate occupations. The variance between départements falls 13%.

TABLE VII.2. - EFFECT OF SELECTED CHARACTERISTICS ON THE PROBABILITY OF STAYING IN THE SAME MUNICIPALITY IN FRANCE

Characteristic	Multilevel logit models				
Fixed	Model 1	Model 2	Model 3	Model 4	Model 5
a_0 (constant)	2.256 (0.042)	2.256 (0.042)	2.149 (0.061)	2.508 (0.067)	2.143 (0.069)
a_1 (woman)	-	-	0.193 (0.053)	0.156 (0.054)	0.120 (0.054)
a_2 (economically inactive)	-	-	-0.280 (0.056)	-0.244 (0.059)	0
a_3 (farmer)	-	-	1.456 (0.327)	1.423 (0.329)	1.271 (0.329)
a_4 (artisan)	-	-	0.657 (0.181)	0.570 (0.182)	0
a_5 (manager)	-	-	-0.223 (0.091)	-0.258 (0.092)	-0.511(0.098)
a_6 (white-collar worker)	-	-	0	0	0
a_7 (manual worker)	-	-	0.473 (0.093)	0.457 (0.094)	0.398 (0.091)
a_8 (unskilled manual worker)	-	-	0.184 (0.110)	0.230 (0.112)	0.346 (0.109)
a_9 (cohabitation)	-	-	-	-1.177 (0.061)	-0.978 (0.062)
a_{10} (marriage)	-	-	-	-0.318 (0.076)	0
a_{11} (separation)	-	-	-	-0.318 (0.060)	-0.271 (0.060)
a_{12} (birth of child)	-	-	-	-0.275 (0.059)	0.346 (0.109)
a_{13} (generation)	-	-	-	-	-1.113 (0.094)
a_{14} (generation^2)	-	-	-	-	0.169 (0.065)
a_{15} (generation^3)	-	-	-	-	0.280 (0.095)
Random variable, level 2	0.084 (0.022)	0.084 (0.022)	0.073 (0.020)	0.073 (0.021)	0.070 (0.020)

Random variable, level 1	1.000	0.988 (0.010)	1.000	1.000	1.000

Model 4 includes the family events occurring in a brief earlier period. We might have assumed that all these events would introduce a strictly local mobility, but they actually introduce a longer-distance mobility. The event that generates the highest mobility is union formation. On the other hand, the introduction of these characteristics has no impact on inter-département variance. These family events have the same effect regardless of the département of residence.

Model 5 introduces a generation effect: as there is no reason for this effect to be linear, we have introduced successive powers of the variable, which we have zero-meaned and divided by 10. The variable strongly influences the behaviors studied; again, however, it has little effect on inter-département variance, which it lessens by only 4%. We know that sedentarity is closely age-related, and the curve obtained by incorporating ages up to the third degree displays the classic profile: it bottoms out at 23 years then declines at a sustained pace at ages 25-35, and more slowly later. Once this effect is introduced, we see that managers and white-collar workers are no longer distinguishable from intermediate occupations. The same is true of the economically inactive: the age structure of these occupations explained the observed mobility differences without calling on this characteristic.

III. - Modeling polytomous data

Let us now suppose that for each individual i present in an area j, we measure a dependent characteristic in the form of a polytomous variable. During the period of study, the individual may experience n categories of events: only one of these events, however, can occur in the period. To model such a situation, it is useful to distinguish between various types of measures of the events.

The events can be purely nominal and therefore unordered: for example, people may or may not change marital status during the observation period. They may move from never-married without previous unions to cohabitant or married, from married to divorced or widowed, etc. They may, of course, remain never-married, cohabiting, married, divorced or widowed during the period. There is no way to order these events.

Other events can, instead, be ordered: for example, if we examine the locality of residence at two dates, the individual may have (1) not moved during the period, (2) moved in the same municipality, (3) changed municipalities but remained in the same département, (3) changed départements but stayed in the same region, or (4) changed regions, if we confine our study to internal migrations in a given country. In this case, the events follow and it may make sense to order them from sedentarity to inter-regional mobility as a function of migration distance; sedentary persons are classified as having traveled a zero distance.

The two situations may not always be easy to distinguish. While the events to be studied can be ordered, each possibility may be determined by different motives, so that we may prefer to model in purely nominal terms. For instance, we may decide that the reasons for migration as a function of distance are very different and unordered: most short-distance migrations may be due to family adjustments, such as marriage or the birth of a child requiring a larger dwelling, without changing one's municipality of residence; longer-distance migrations may be due to an occupational change demanded by the employer or to finding a far more rewarding job in another region. In this case, it is preferable to introduce unordered events, so

as to analyze mobility as a function of characteristics exerting a strong influence on one type of event and a weak influence on another type. Later on, we shall examine the underlying hypotheses of the models used in both cases, which can be treated separately.

Let us note here that other, more complex events enter into this type of analysis. Suppose, for example, that we want to analyze simultaneously the fact that an individual smokes or not (binary characteristic), during a given period, and, if he or she smokes, the average number of cigarettes smoked daily (a characteristic that we may regard as continuous). This becomes a mixed model simultaneously involving a binary dependent characteristic and another, continuous characteristic.

Explanatory characteristics may themselves be dichotomous, polytomous or continuous, as in models explaining binary data. We may regard them as fixed characteristics or as random characteristics operating at some aggregation levels, as described later.

We begin with a concise theoretical description of the different types of models with applications to observed data, in order to show their advantages and disadvantages.

Models to explain nominal characteristics

As an example, we take the case where the number of categories is equal to n. We assume there is only one explanatory characteristic, x_{1ij}, and n random characteristics, u_{0j}^{1}, ... ,u_{0j}^{n}, for each category, operating at the aggregate level.

We suppose that a multivariate logit model applies to these data correctly. The probability that individual i present in area j displays the k[th] category—in which (k = 1, …. , n)—can be written:

$$P(y_{ij} = k \mid x_{1ij}) = \pi_{ij}^{k} \qquad \text{(VII.19)}$$

We can generalize the previous logit model (VII.12) by choosing one category, for example the last, as the benchmark category in order to avoid redundancies and a non-invertible covariance matrix. The model is written:

$$\pi_{ij}^{k} = \exp(a_0^k + a_1^k x_{1ij} + u_{0j}^k)\left[1 + \sum_{h=1}^{n-1} \exp\left(a_0^h + a_1^h x_{1ij} + u_{0j}^h\right)\right]^{-1} \qquad \text{(VII.20)}$$

The effect of the characteristic x_{1ij} on the probabilities for each category yields the following multiplicative factors:

1, $\exp(a_1^1 x_{1ik})$, ... ,$\exp(a_1^{n-1} x_{1ik})$

which show how to shift from the effect of the first category to that of the second and so forth.

As with the binary data, we can also replace this multivariate logit model by a multivariate log-log model, generalizing the log-log model presented in part II of this chapter. We can accordingly write:

$$\pi_{ij}^{k} = \left[1 - \exp\left(-\exp(a_0^k + a_1^k x_{1ij} + u_{0j}^k)\right)\right] \qquad \text{(VII.21)}$$

As this function does not involve the benchmark category, the choice of the latter now becomes important and can alter the results of the analysis significantly. This option must therefore be used with caution.

Going back to the logit model, we see that the covariance matrix at the aggregate level can be written:

$$\begin{pmatrix} \pi_{ij}^{1}(1-\pi_{ij}^{1}) & & & \\ -\pi_{ij}^{1}\pi_{ij}^{2} & \pi_{ij}^{2}(1-\pi_{ij}^{2}) & & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot \\ -\pi_{ij}^{1}\pi_{ij}^{n-1} & -\pi_{ij}^{2}\pi_{ij}^{n-1} & \cdot \quad \cdot & \pi_{ij}^{n-1}(1-\pi_{ij}^{n-1}) \end{pmatrix} \qquad \text{(VII.22)}$$

As above, we obtain another non-linear model, which we can estimate using the general procedure described in part I of this chapter.

Let us illustrate the application of this model to concrete data through one specific example.

Example no. 8 (continued): Migrants in year before "Young People and Careers" survey in France

Going back to the previous example, which we had analyzed with the aid of a logit model, we now divide the sedentary—as defined earlier—into two categories: (1) individuals who have performed no migration during the year, and (2) individuals who have performed one intra-municipal migration. We naturally continue to separate these categories from individuals migrating outside their municipality of residence. This will enable us to see whether or not intra-municipal migrants can be distinguished from truly sedentary individuals.

Given that the control group defined in this model is arbitrary, we can—as in the previous example— choose inter-municipal migrants as such, bearing in mind that we can determine any group's probabilities of migrating from these estimations.

Model 1 enables us to estimate the probabilities of the three types of events that we can calculate with the aid of two parameters estimated for the fixed characteristics. The probability of not migrating is equal to:

$$P(m_0) = \frac{e^{2.162}}{1 + e^{2.162} + e^{-0.422}} = 0.8399$$ versus 0.8343 calculated on raw data,

that of migrating within a municipality:

$$P(m_i) = \frac{e^{-0.422}}{1 + e^{2.162} + e^{-0.422}} = 0.0634$$ versus 0.0668 calculated on raw data,

and that of migrating outside a municipality:

$$P(m_e) = \frac{1}{1 + e^{2.162} + e^{-0.422}} = 0.0967$$ versus 0.0989 calculated on raw data,

but these raw estimates do not take account of the file's hierarchical structure. We do confirm that the sum of these probabilities is equal to unity.

We can also adopt the same approach to estimate the same probabilities for each département using the level-2 random terms, which are totally significant. Figure VII.1 reports the results.

TABLE VII.3. - EFFECT OF SELECTED CHARACTERISTICS ON NON-MIGRANTS AND INTRA-MUNICIPAL MIGRANTS IN FRANCE

Characteristics	Model 1		Model no. 2	
Fixed	Non-migrants	Intra-municipal migrants	Non-migrants	Intra-municipal migrants
a_0 (constant)	2.162 (0.040)	-0.422 (0.064)	2.003 (0.061)	-0.753 (0.087)
a_1 (woman)	-	-	0.189 (0.034)	0
a_2 (economically inactive)	-	-	-0.242 (0.059)	0.337 (0.079)
a_3 (farmer)	-	-	1.505 (0.209)	1.114 (0.261)
a_4 (artisan)	-	-	0.731 (0.123)	0
a_5 (manager)	-	-	-0.162 (0.072)	0
a_6 (white-collar worker)	-	-	0.128 (0.065)	0.440 (0.088)
a_7 (manual worker)	-	-	0.522 (0.078)	0.439 (0.108)
a_8 (unskilled manual worker)	-	-	0.198 (0.088)	0.681 (0.117)
Random variables, level 2:	0.099 (0.021) 0.073 (0.024) 0.261 (0.054)		0.096 (0.021) 0.077 (0.024) 0.272 (0.056)	
σ^2_{u0s} (non-migrants)				
σ_{u0sm} (covariance)				
σ^2_{u0m} (migrants)				
Random variable, level 1	1		1	

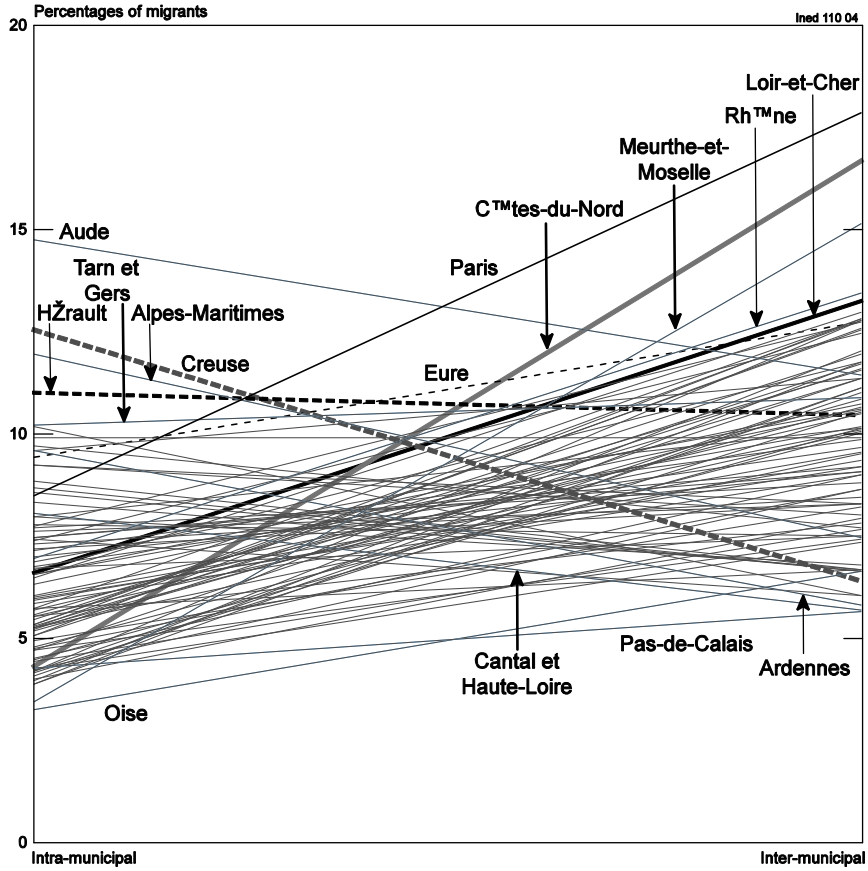


Figure VII.1. - Proportions of intra- and inter-municipal migrants estimated with multiple-choice model

This figure shows that, while inter-municipal migration outweighs intra-municipal migration in a majority of départements, some départements do not fit into the general pattern: Alpes-Maritimes, Aude, Creuse, Ardennes, Cantal, Haute-Loire, and, to a lesser extent, Hérault display a higher intra-municipal mobility. By contrast, some strongly urbanized départements have a very high inter-municipal mobility: Paris, Rhône, Meurthe-et-Moselle, Oise, and, to a lesser extent, Pas-de-Calais. The raw results estimated for each département separately yield similar estimates. As a rule, however, there are no significant differences between them, given the small number of individuals observed. The advantage of the multilevel model is that it takes into account the number of individuals observed and ensures the significance of the differences between random variables.

At the individual level, we can introduce a random variable whose variance is not necessarily equal to unity as before. We have not reported this model in table VII.3, for it shows that the hypothesis of a multivariate logit model is still properly confirmed: the level-1 random variable obtained does not always differ significantly from unity.

The second model includes the individual's sex and occupation a year before the survey. If the effect of occupation was restricted to intra-municipal migrations, the parameters estimated for non-migrants and intra-municipal migrants should not be significantly different. We find this pattern for farmers and manual workers, but it does not apply to other occupations. For instance, the behavior of artisans and managers is identical for intra- and inter-municipal migrants and different for sedentary individuals. The same is also true of women's behavior. White-collar workers and unskilled manual workers display higher intra-municipal mobility than inter-municipal mobility: their labor market is narrower than that of managers. The behavior of the economically inactive is even more complex: relative to the

overall model indicated by the constant term, non-migrants form the smallest group, intra-municipal migrants the largest group, and inter-municipal migrants fall in between.

By contrast, we note that the random variables at the département level are practically unchanged, which means that figure VII.1 continues to apply—with the exception of one overall translation—to the behaviors of the various occupations and to that of women.

Table VII.4 extends this analysis by introducing family events and the generation effect.

TABLE VII.4. - EFFECT OF SELECTED CHARACTERISTICS ON NON-MIGRANTS AND INTRA-MUNICIPAL MIGRANTS IN FRANCE (CONTINUED)

Characteristics	Model 3		Model 4	
Fixed	Non-migrants	Intra-municipal migrants	Non-migrants	Intra-municipal migrants
a_0 (constant)	2.403 (0.066)	-0.828 (0.095)	2.229 (0.069)	-0.830 (0.096)
a_1 (woman)	0.153 (0.035)	0	0.109 (0.036)	0
a_2 (economically inactive)	-0.218 (0.061)	0.369 (0.083)	-0.173 (0.065)	0.438 (0.088)
a_3 (farmer)	1.459 (0.212)	1.116 (0.265)	1.367 (0.214)	1.071 (0.267)
a_4 (artisan)	0.645 (0.145)	0	0.397 (0.147)	0
a_5 (manager)	-0.198 (0.075)	0	-0.361 (0.076)	0
a_6 (white-collar worker)	0.133 (0.067)	0.473 (0.092)	0.239 (0.069)	0.488 (0.093)
a_7 (manual worker)	0.500 (0.080)	0.462 (0.111)	0.534 (0.082)	0.479 (0.111)
a_8 (unskilled manual worker)	0.265 (0.091)	0.746 (0.120)	0.478 (0.093)	0.783 (0.122)
a_9 (cohabitation)	-1.313 (0.052)	-0.271 (0.072)	-1.095 (0.054)	-0.230 (0.073)
a_{10} (marriage)	-0.393 (0.061)	0.290 (0.081)	-0.134 (0.062)	0.341 (0.081)
a_{11} (separation)	-0.352 (0.039)	0	-0.298 (0.040)	0
a_{12} (birth of child)	-0.340 (0.047)	0.230 (0.064)	-0.152 (0.048)	0.251 (0.064)
a_{13} (generation)	-	-	-1.101 (0.063)	-0.154 (0.049)
a_{14} (generation^2)	-	-	0.113 (0.040)	0
a_{15} (generation^3)	-	-	0	0.231 (0.060)
Random variables, level 2:	0.101 (0.022)		0.099 (0.022)	
σ^2_{u0s} (non-migrants)	0.088 (0.026)		0.088 (0.026)	

σ_{u0sm} (covariance)	0.294 (0.059)	0.298 (0.060)
σ_{u0m} (migrants)		
Random variable, level 1:	1	1

Model 3 includes, in addition to the previous characteristics, the family changes experienced by individuals in the two years prior to migration. If these events generated only intra-municipal mobility, we would observe parameters significantly different from zero for intra-municipal migrants; sedentary persons and inter-municipal migrants would not be affected. Once more, the results fail to confirm this hypothesis. While we do observe an increase in the probability of intra-municipal migration by comparison with the probabilities of non-migration and inter-municipal migration—except in the event of separation—the probability of inter-municipal migration because of these events is greater than the probability of not migrating. The introduction of these family characteristics entails almost no change in the effect of occupations and especially in the random variables at the département level, which remain identical to what they were in the previous models.

Model 4 introduces the effect of the individual's generation. This characteristic has different effects on non-migrants and intra-municipal migrants, but does not significantly modify the effects demonstrated earlier. In particular, it leaves the random variables at the département level unchanged. It is interesting to observe this constancy, whereas the logit model applied earlier to similar data entailed a 17% reduction of this random variable when we incorporated all of the same characteristics.

Models to explain ordinal characteristics

Let us continue to assume that the variable has only three categories, but that we can arrange them in a particular order. Let us further assume that the choice of a given category is independent of the characteristics influencing the phenomenon studied, and that it introduces only a constant term, which depends on the category itself. This situation leads to models that are based not on each category, as before, but on the cumulative distribution of the categories up to the kth (McCullagh and Nelder, 1989):

$$E(y_{ij}^{k}) = \gamma_{ij}^{k} = \left[1 + \exp\left\{-(a_0^k + a_1 x_{1ij} + u_{0j})\right\}\right]^{-1} \quad \text{where k = 1,…, n-1} \qquad \text{(VII.23)}$$

and n is the total number of categories.

This assumes, of course, that $a_0^1 \le a_0^2 \le \ . \ . \ . \le a_0^{n-1}$, since we are working on cumulative distributions. Again, we see that the a_1 parameter measures the effect of the characteristic x irrespective of the category observed, and that the random variable at the aggregate level is independent of the category. We can generalize this model by introducing random variables that also depend on the category (Fielding et al., 2003). The model is then written:

$$E(y_{ij}^{k}) = \gamma_{ij}^{k} = \left[1 + \exp\left\{-(a_0^k + a_1 x_{1ij} + u_{0j}^{k})\right\}\right]^{-1} \qquad \text{(VII.24)}$$

This model can easily be generalized to any given number of explanatory characteristics x, it can incorporate random variables for a larger number of characteristics (usually for some of the characteristics x introduced) or it can accommodate more than three categories.

Once again, we can estimate these non-linear models using methods similar to those presented in VII.1.

Let us now see the results of the application of this type of model to the French migration data.

Example no. 8 (continued): Migrants in year before "Young People and Careers" survey in France

In chapter VI, we saw that the conditions for an effective application of a simple ordinal model of the VII.23 type are far from satisfied. The random variables given in figure VII.1 are anything but ordered, although we can discern a rough trend: the probability of an inter-municipal migration is generally greater than that of an intra-municipal migration. In a second phase, we shall therefore need to apply model VII.24, which enables us to vary the model's random variables with the département. Table VII.5 shows the results of such an analysis, which will need to be compared with those of tables VII.2, VII.3, and VII.4.

The first model includes only the cumulative probabilities: first the probability of being sedentary, then the probability of being both sedentary and an intra-municipal migrant. As in the previous model, we can determine the elementary probabilities from these figures. For instance, we can write the probability of not migrating as:

$$P(m_0) = [1 + \exp(-1.656)]^{-1} = 0.8397$$

that of intra-municipal migration as:

$$P(m_i) = [1 + \exp(-2.256)]^{-1} - P(m_0) = 0.0655$$

and that of inter-municipal migration as:

$$P(m_e) = 1 - [1 + \exp(-2.256)]^{-1} = 0.0948$$

Comparing these figures with the ones given in the previous example, we find very similar values, which do not allow us to conclude that one model is superior to the other. The level-2 random variables will be the main cause of incorrect estimates for each département, as figure VII.1 shows that the model's hypotheses are not confirmed. To verify this, let us take some départements and compare the estimates supplied by this model, the model with nominal characteristics, and the observations in table VII.6.

TABLE VII.5 - EFFECT OF SELECTED CHARACTERISTICS ON NON-MIGRANTS AND INTRA-MUNICIPAL MIGRANTS IN FRANCE

Characteristics	Multilevel model with ordinal characteristics				
Fixed	Model 1	Model 2	Model 3	Model 4	Model 5
a_{01} (non-migrant)	1.656 (0.038)	1.658 (0.038)	1.664 (0.039)	1.614 (0.039)	1.876 (0.060)
a_{02} (intra-municipal migrant)	2.256 (0.040)	2.256 (0.040)	2.231 (0.041)	2.180 (0.041)	2.493 (0.062)
a_1 (woman)	-	-	-	-	0.106 (0.044)
a_2 (inactive)	-	-	-	-	0
a_3 (farmer)	-	-	-	-	0.906 (0.215)
a_4 (artisan)	-	-	-	-	0.233 (0.142)
a_5 (manager)	-	-	-	-	0.407 (0.078)
a_6 (white-collar worker)	-	-	-	-	0
a_7 (manual worker)	-	-	-	0.395 (0.064)	0.346 (0.072)
a_8 (unskilled manual worker)	-	-	-	-	0.140 (0.082)
a_9 (cohabitation)	-	-	-	-	0.996 (0.053)
a_{10} (marriage)	-	-	-	-	0.239 (0.062)
a_{11} (separation)	-	-	-	-	0.307 (0.049)
a_{12} (birth of child)	-	-	-	-	0.232 (0.048)
a_{13} (generation)	-	-	-	-	1.033 (0.076)
a_{14} (generation^2)	-	-	-	-	0.141 (0.051)
a_{15} (generation^3)	-	-	-	-	0.209 (0.076)
Random variable, level 2				-	
σ^2_{u01} (non-migrant)	0.078 (0.018)	0.078 (0.018)	0.087 (0.019)	0.084 (0.019)	0.076 (0.018)
σ_{u12} (covariance)			0.059 (0.017)	0.054 (0.017)	0.045 (0.017)
σ^2_{u02} (intra-municipal migrant)			0.082 (0.022)	0.080 (0.021)	0.076 (0.021)
Random variable, level 1	1	0.991 (0.007)	1	1	1

TABLE VII.6. - COMPARISON OF PROBABILITIES SUPPLIED BY PROPORTIONAL-EFFECTS MODEL AND SEPARATE-EFFECTS MODEL AGAINST RAW OBSERVATION DATA

Models:	Alpes-Maritimes		Aude		Bouches-du-Rhône		Paris	
	Intra-municipal migration	Inter-municipal migration	Intra-municipal migration	Inter-municipal migration	Intra-municipal migration	Inter-municipal migration	Intra-municipal migration	Inter-municipal migration
Proportional effects	0.0750	0.1122	0.0984	0.1619	0.0697	0.1024	0.1006	0.1672
Separate effects	0.1195	0.0750	0.1478	0.1144	0.0926	0.0831	0.0851	0.1788
Observations	0.1325	0.0728	0.2742	0.1452	0.0977	0.0818	0.0909	0.1864
Number of persons	302		62		440		660	

We see that for the Alpes-Maritimes, where observed intra-municipal migrations outweigh inter-municipal migrations, the estimate obtained with the proportional-effects model is completely aberrant. For the Aude, where the number of individuals observed is small, the two estimates diverge rather substantially from the observations, but the separate-effects model does supply an estimate closer to the observed values. For the Bouches-du-Rhône, the estimation with separate effects is almost perfect. For Paris—where the hypothesis that inter-municipal migrations outweigh intra-municipal migration is confirmed—the three estimates converge closely: however, even here, the separate-effects model yields a better estimate than the proportional-effects model. The examination of the other residuals confirms that the proportional-effects model is ill-suited to the phenomenon studied.

If we now free the variance of the random variable at the individual level, model 2 shows that this random variable remains not significantly different from unity. A multivariate logit model is therefore always well-suited to these data, although the random variable at the département level is consistently ill-suited. We therefore cannot detect the poor quality of the département random variable by allowing the random variable at the individual level to vary. We need to incorporate a variation of the département random variable according to the type of mobility considered.

Model 3 introduced this random variable for non-migrants and another for intra-municipal migrants. These random variables are entirely significant and enable us to reconstruct a figure equivalent to figure VII.1, for a model with ordinal characteristics, by using the residuals to compute the proportions of intra-municipal migrants and inter-municipal migrants. We can write:

$$P(m_0) = \left[1 + \exp -(a_{01} + u_{01})\right]^{-1}$$

$$P(m_i) = \left[1 + \exp -(a_{02} + u_{02})\right]^{-1} - P(m_0)$$

$$P(m_e) = 1 - \left[1 + \exp -(a_{02} + u_{02})\right]^{-1}$$

Figure VII.2 plots these random variables at the département level. We can assess the quality of this model against that of the previous model by comparing their results with the observed data using a χ^2 test. Thus, if we observe m_{ij} intra-municipal migrants in département j, whose population is p_j, then the sum of the squares of Pearson residuals, $\frac{m_{ij} - p_j P(m_{ij})}{\sqrt{p_j P(m_{ij})[1-P(m_{ij})]}}$, leads to this test. Such a statistic yields a better estimate of the flows of intra-département migrants by the ordered model (34.03 versus 68.45 for 83 degrees of freedom) and an inferior estimate for inter-département flows (37.01 versus 20.69). We can therefore conclude that both models provide good estimates of these flows.

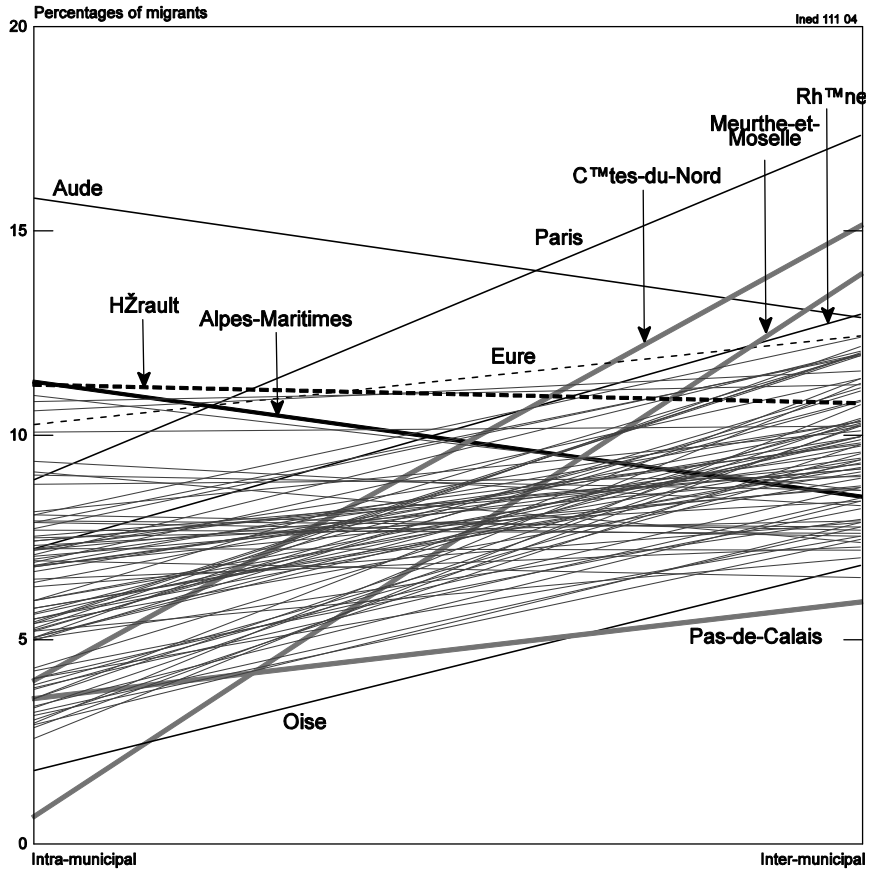


Figure VII.2. - Proportions of intra- and inter-municipal migrants estimated using the ordinal model with département-dependent random variables

If we now try to include the effect of various characteristics, we have seen that it was sometimes consistent with an ordered model, and sometimes not. We must therefore begin by observing what happens if we use a variable whose effect seems to validate a model with ordinal characteristics. For this purpose, let us take the case of manual workers. Before applying the multilevel model, let us see in greater detail how to illustrate the application of a model without a département level. Table VII.7 lists the raw numbers of individuals observed in the survey and allows a preliminary search to determine whether an ordered model can be applied properly.

TABLE VII.7. - NUMBER OF FARMERS AND NON-FARMERS BY MIGRATION STATUS

	Non-migrants	Intra-municipal migrants	Inter-municipal migrants	Total
Manual workers	2,318	140	176	2,634
Non-manual workers	13,673	1,140	1,720	16,533
Total	15,991	1,280	1,896	

The logit of the first contrast to be examined for non-migrant manual workers (2,318 versus 140 + 176) yields the value log (2,318.5 / 316.5) = 1.991, as it is preferable to add 0.5 to the fraction's numerator and denominator in order to reduce biases and avoid zero values (McCullagh, 1980). The second contrast also concerns the combined set of non-migrants and intra-municipal migrants, again among manual workers; it is similarly written as log (2,485.5 / 176.5) = 2.640. A comparable calculation for non-manual workers yields the opposing values of 1.564 and 2.153. The difference between the non-migrant contrasts is 0.427; the difference between the contrasts for non-migrants and intra-municipal migrants combined equals 0.481. From the relative proximity between these two figures and their identical sign, we can deduce the validity of an ordered model based on these data. The application of an ordered logit model to these figures shows a 64% decrease in manual workers' probability of inter-municipal migration compared with other occupations, and an identical decrease in the probability of remaining sedentary for other occupations compared with manual workers.

Multilevel model 4 includes the manual-worker characteristic and yields fixed parameters identical to those of the simple ordinal logit model described in the previous section: 1.614 versus 1.564 for non-migrants, 2.180 versus 2.158 for intra-municipal migrants, and 0.395 versus 0.432 for manual workers. There is little change in the random variables at département level by comparison with multilevel model 3, which did not incorporate manual workers.

However, once we include the full set of characteristics examined here in the last multilevel model (no. 5), the model, while still supplying correct results in certain cases, gives far less satisfactory results in others. For farmers or for persons who have just started cohabiting, the ordinal model clearly continues to produce results close to those of the nominal model. For women, artisans, managers, and also individuals experiencing a separation, we see that the ordinal-model results diverge from the nominal-model results, as the nominal model shows that we cannot distinguish the behavior of intra-municipal migrants from that of inter-municipal migrants. Some individuals with specific characteristics no longer even conform to an ordinal model at all: the economically inactive, for example, have a zero parameter in the ordered model whereas the nominal model indicates a more complex behavior, as noted in our earlier analysis of the model.

We can therefore conclude that, when applied to French migrations, the ordinal model functions well for certain characteristics but yields incorrect results for others. Indeed, if we try selecting non-migrants as the dummy category, the model no longer converges. Before undertaking such a study, it is therefore essential to check whether the model conditions are properly satisfied. If they are not, we must use a formalization where the various categories act separately or a two-stage model, which examines (1) whether the individual is a migrant or not,

and (2) whether he or she is an intra- or inter-municipal migrant (see, for example, the model elaborated by Craig Duncan (1997), in which he begins by separating non-smokers from smokers, then classifies smokers by their daily consumption).

These models can be applied correctly to biometric data for which individuals belonging to different groups can be classified into a given number of ordered categories: the characteristics measured simultaneously will have a general influence on the various measures, each of these introducing a specific overall effect.

Other applications of the models are found in education science, when students are classified into categories (Fielding et al., 2003)

Modeling an event count

Let us turn now to event counts: the number of migrations performed between ages 16 and 40, for example, or the number of children to whom a women has given birth during her fertile life. Of course, we can always model this count as a continuous variable, especially when the number of events is high and more or less normally distributed across the spectrum. As this number is positive—by contrast with the Normal distribution, which covers the entire set of real numbers—it is preferable here to model the logarithm of the number of events: we obtain the models discussed in chapter VI.

By contrast, when the number of events is small, and when their distribution or that of their log differs substantially from a normal law, it is better to use a model with discrete characteristics to study the effect of particular characteristics on the number of events.

Let us suppose that, for each individual, we have the number of events experienced during a given fixed or variable period. For example, we know the number of sick leaves taken by an individual in a given year or time interval, which may vary from one individual to another. Let y_{ij} be this number measured for an individual i, present in the enterprise or area j during the period considered. When this period is of variable length, t_{ij}, it is useful to include it in the mode, for the number of sick leaves should increase when the period lengthens. Such a variable functions as an "output"—i.e., an adjustment of the characteristic observed. It may, of course, differ from a time interval: for example, if we study the number of sick leaves in a given year, we can examine the number of sick leaves in the previous year. Lastly, let us assume that we know one of the individual's characteristics—either a continuous one, such as age, or a discrete one, such as sex or occupation, x_{1ij} —that should influence the number of events studied. As usual, of course, we can generalize this model to any number of individual or aggregate characteristics.

We must now consider various modelings of positive variables used in statistics. The classic model is the Poisson model, in which events occur completely at random over time. The dependence between $E(y_{ij})$, called here π_{ij}, and the characteristics is assumed to be multiplicative, leading to the following log-linear multilevel model:

$$\log\left\lfloor E(y_{ij})\right\rfloor = \log(\pi_{ij}) = \log(t_{ij}) + a_0 + a_1 x_{1ij} + u_{0j} \quad \text{(VII.25)}$$

where the term $\log(t_{ij})$ may be present or not. We can write the variance of the observations as:

$$\text{var}(y_{ij}\left|\pi_{ij}\right.) = \pi_{ij}$$

(VII.26)

Again, we end up with a non-linear model whose parameters we need to estimate using the methods briefly described in VII.1.

However, there is a possibility—especially in social science—that this model, in which the mean and the variance of the number of events are identical, may not be confirmed. Statistically speaking, this phenomenon can occur in different ways that it is not useful to describe in detail here (see, for example, McCullagh and Nelder, 1989). This leads us to estimate a variance of the number of events proportional to its mean:

$$\text{var}(y_{ij}\left|\pi_{ij}\right.) = k_1\pi_{ij}$$

(VII.27)

The result is thus a model in which the variance of the number of events is proportional to its mean: the log-linear model continues to apply.

Alternatively, the variance of the number of events may be proportional not to its mean, but to the square of the mean. If so, the constant term is the variation coefficient. We use a Gamma distribution, which satisfies the relationship:

$$\text{var}(y_{ij}\left|\pi_{ij}\right.) = k_2\pi_{ij}^2$$

(VII.28)

We may also be faced with a phenomenon that obeys a Poisson process, but whose mean follows a Gamma random distribution. The result is a negative binomial distribution, which leads us to estimate a variance of the form:

$$\text{var}(y_{ij}\left|\pi_{ij}\right.) = \pi_{ij} + k_2\pi_{ij}^2$$

(VII.29)

Lastly, we may examine an even more general distribution where the first-degree term is also a function of a parameter to be estimated:

$$\text{var}(y_{ij}\left|\pi_{ij}\right.) = k_1\pi_{ij} + k_2\pi_{ij}^2$$

(VII.30)

In sum, we may encounter all the previous variances for different values of the k_1, k_2 parameters. Of course, we can test which is the best distribution for the phenomenon studied and, when none of the earlier distributions is suitable, we can always use more complex ones. However, we already have here a wide range of distributions to choose from.

Example no. 9: Number of migrations performed from age 16 up in "Young People and Careers" survey in France

For this example, we take the generations born between 1952 and 1958 to obtain a sufficiently large population. Our adjustment variable will thus be the time elapsed between the respondent's 16th birthday and the survey date: the mean number of migrations by the first generation observed is 5.76 versus only 5.24 for the youngest. We shall attempt to explain this number of migrations by various characteristics measured at the start of the individual's working life: school-leaving age, sex, first occupation, region, and category of municipality of residence. We use a regional segmentation of France, which, however, breaks down the most populated regions into three areas for Île-de-France (Paris, inner ring of suburbs, and outer ring of suburbs) and into two areas for Rhône-Alpes (Rhône + Haute-Savoie, rest of the region).

Let us begin by determining which type of model will work best with these data. Table VII.8 reports how the models described earlier apply to the simplest specification, which includes the mean number of migrations and random variables at the regional and individual levels.

TABLE VII.8. - SEARCH FOR BEST TYPE OF RANDOM VARIABLE TO INCLUDE FOR NUMBER OF MIGRATIONS

Characteristics	Simple multilevel models			
Fixed	Model 1	Model 2	Model 3	Model 4
a_0 (constant)	-1.616 (0.029)	-1.616 (0.029)	-1.615 (0.029)	-1.614 (0.028)
Random variable, level 2: σ^2_{u0}	0.020 (0.006)	0.018 (0.006)	0.018 (0.006)	0.017 (0.006)
Random variable, level 1: σ^2_{e0}	1	2.885 (0.057)	1	0.000 (0.000)
σ^2_{e1}	-	-	0.342 (0.011)	0.5192 (0.010)

The first model is a pure Poisson model, with a unit variance of the individual-level random variable. If we let this variance vary freely (model 2), its value greatly exceeds unity, totally disqualifying the Poisson model. We should therefore introduce a second term in this variance at the individual level. Model 3 is a negative binomial model, where the first parameter's variance is equal to unity: the second parameter is significantly different from zero. We can go further by allowing the first parameter to vary freely (model 4). This gives us a Gamma-model variance, since it is proportional to the square of π_{ij}, and the first-degree term cancels out. Later on, therefore, we shall apply a model of this type to our data.

Table VII.9 reports the model's application to the characteristics examined here. Let us begin with the effect of school-leaving age on the mean number of dwellings occupied between ages 15 and 42.

TABLE VII.9. - EFFECT OF SELECTED CHARACTERISTICS ON NUMBER OF DWELLINGS OCCUPIED BETWEEN AGES 15 AND 42

Characteristics	Gamma model			Poisson model
Fixed	Model 1	Model 2	Model 3	Model 1
a_0 (constant)	-1.729 (0.042)	-1.699 (0.044)	-1.570 (0.057)	-1.595 (0.049)
a_1 (school-leaving age)	0.297 (0.050)	0.298 (0.050)	0.203 (0.050)	0.178 (0.054)
a_2 (woman)	-	-0.049 (0.020)	-0.074 (0.023)	-0.067 (0.014)
a_3 (only child)	-	-0.076 (0.038)	-0.082 (0.038)	-0.091 (0.024)
a_4 (1 sibling)	-	0	0	-0.037 (0.015)
a_5 (farmer)	-	-	-0.378 (0.065)	-0.348 (0.049)
a_6 (artisan)	-	-	-0.258 (0.072)	-0.229 (0.047)
a_7 (manager)	-	-	0	0.085 (0.029)
a_8 (white-collar worker)	-	-	-0.067 (0.033)	-0.051 (0.021)
a_9 (manual worker)	-	-	-0.095 (0.040)	-0.073 (0.024)
a_{10} (unskilled manual worker)	-	-	-0.134 (0.037)	-0.110 (0.023)
a_{11} (Paris area)	-	-	0	0.089 (0.039)
a_{12} (big cities)	-	-	0.072 (0.035)	0.080 (0.022)
a_{13} (medium-sized localities)	-	-	0	0.033 (0.015)
Random variables, level 2:				
σ^2_{u0}	0.038 (0.013)	0.038 (0.013)	0.041 (0.013)	0.043 (0.013)
σ_{u01}	-0.034 (0.014)	-0.035 (0.014)	-0.037 (0.013)	-0.042 (0.014)
σ^2_{u1}	0.031 (0.017)	0.033 (0.017)	0.030 (0.015)	0.053 (0.018)
σ_{u08}	-		-0.011 (0.004)	0
Random variables, level 1:				
σ^2_{e0}	0	0	0	1
σ^2_{e1}	0.503 (0.010)	0.506 (0.010)	0.505 (0.010)	-

Model 1 introduced school-leaving age from age 15 up (divided here by 10), at both the fixed and regional levels. The number of dwellings occupied since age 15 generally rises as a function of school-leaving age. But the examination of the regional curves obliges us to strongly qualify this conclusion. Figure VII.3 plots these residuals calculated as the exponential of the model's raw result multiplied by the mean observation period, here 27 years. This effectively yields exponential curves giving the mean number of dwellings occupied by the generation born in 1956, by school-leaving age and region of initial residence. It is interesting to compare these results with those of a simple Poisson model (figure VII.4) and a simple linear-regression model (figure VII.5).

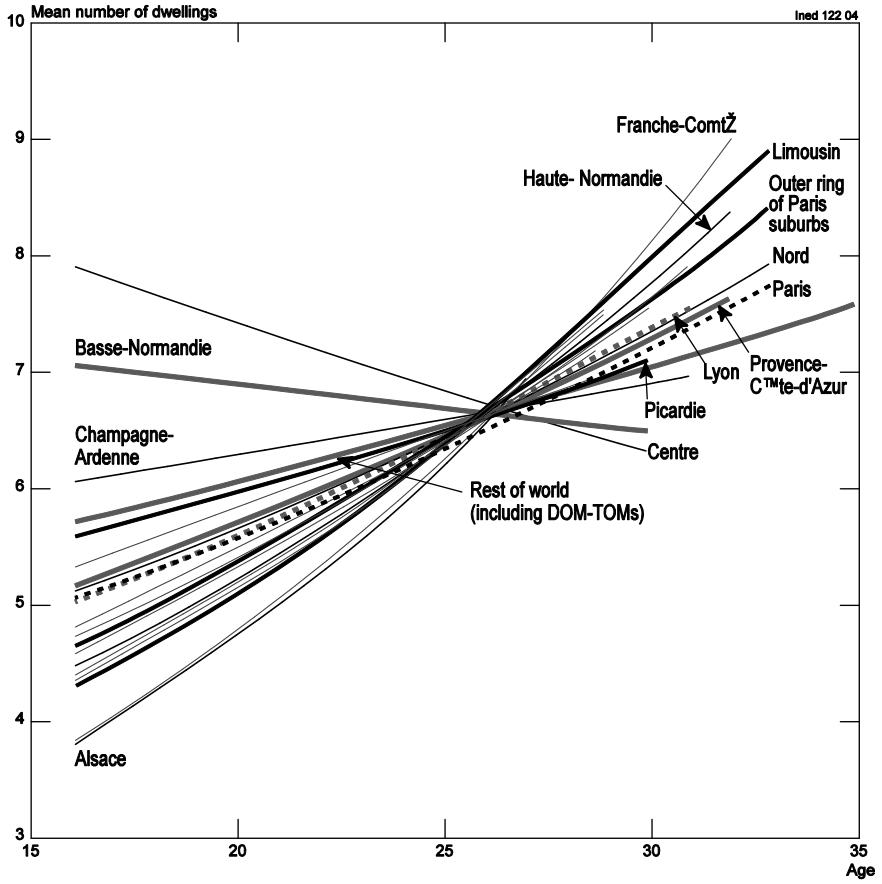


Figure VII.3. - Mean number of dwellings by age (negative binomial model)

Figure VII.3 shows curves that rise with school-leaving age except in two regions: Centre and Basse-Normandie. The regions displaying the steepest increases as a function of school-leaving age are Alsace, Franche-Comté, and Limousin. All the curves more or less intersect a common point located at about 26 years, so that the regional ranking is completely reversed beyond that age. The curves obtained with a simple Poisson model—which therefore captures the individual-level random variables less effectively—are far more scattered (figure VII.4). The curve for Paris, for example, is practically parallel to the x-axis, whereas the model that takes into account the individual random variables correctly gives a far stronger variation as a function of school-leaving age. We do, however, find the previous decreasing curves for the Centre and Basse-Normandie, and the sharp increases for Alsace, Franche-Comté, and Limousin. One region stands apart for its steep rises as well: the Lyon region, which the previous model placed in the regional mean. But, most importantly, these curves no longer

have any common intersection point whatsoever, making this figure hard to compare with the previous one. If we now apply a classic regression model to the number of migrations treated as continuous variables, disregarding the effect of the auxiliary variable (the observation period), we obtain figure VII.5. Its results are far closer to the optimal model than to the simple Poisson model. The regions' situation is broadly similar and the existence of an intersection point for the curves is more visible, albeit less distinctly than in the optimal model. The results of a given region are now perfectly aligned, because of the model used, but this does not introduce a major difference with the previous estimates, in which these results lay on an exponential curve. Here, the use of a linear regression seems preferable to a simple Poisson model, although less satisfactory than the model initially described.

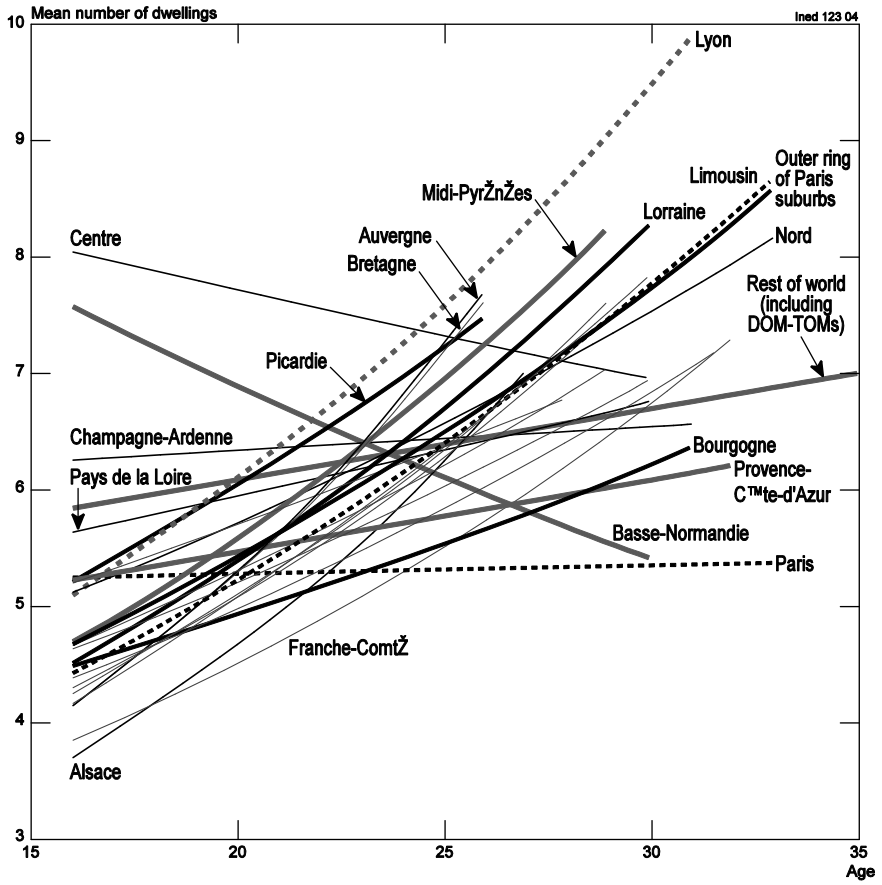


Figure VII.4. - Mean number of dwellings by age (Poisson model)

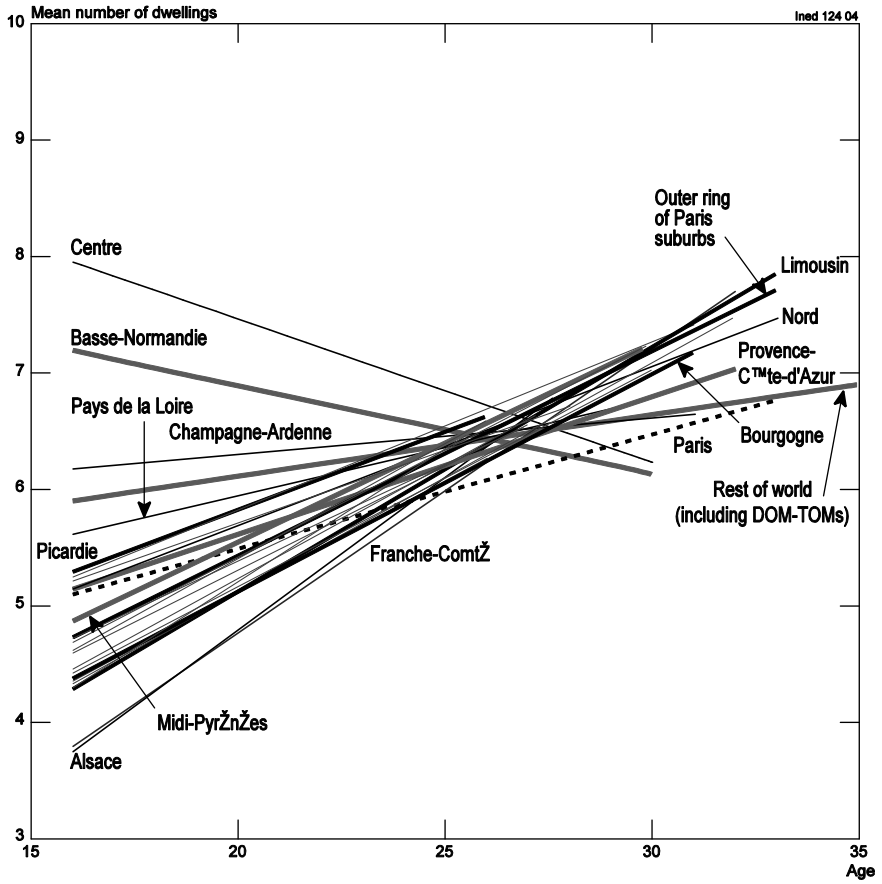


Figure VII.5. - Mean number of dwellings by age (normal model)

Model 2 incorporates two new characteristics: the respondent's sex and number of siblings. Women report fewer dwelling changes, as already determined: in fact, it is before the start of a union that men exhibit higher mobility than women, but we cannot confirm the finding with this file. For the number of siblings, rather than including the characteristic in its pure form, we preferred to replace it by a series of dichotomous variables for each number. Actually, it seems that only children are the only category to stand apart from the others with a lower mobility—doubtless due to the fact that they more commonly inherit their parent's property. The level-1 and level-2 random variables remain identical to their values in the previous model.

Model 3 includes all the characteristics considered here. The effects of the characteristics examined earlier remain unchanged. The ranking of occupations by mobility, in decreasing order from the highest to the lowest, is as follows: managers, white-collar workers, skilled manual workers, unskilled manual workers, artisans, and farmers. If we look at the sizes of the localities of residence for persons living with their parents, the only major distinction concerns individuals residing in large cities, who display a higher mobility.

It is interesting to compare the results provided by this model with the ones obtained from a simple Poisson model: the results are shown in the last column of table VII.9. We have already noted the difference in the level-2 random variables when the only characteristic included was school-leaving age: these differences remain significant. Our attention will focus, instead, on the effect of fixed characteristics. While the effects exhibit no sign change, subtler differences appear. Owing to a poor specification of the random variable at the individual level, the variances of the estimated fixed parameters will generally be smaller than those given by the previous model. The introduction of an incorrect formulation of the individual-level variance will bias the confidence intervals of the fixed effects by narrowing them significantly,

just as the omission of a regional level could introduce a similar reduction of the variance of the fixed effects (see chapter VI). In consequence, characteristics that had no significant effect in the previous model acquire one in this model: presence of siblings, childhood residence in the Paris agglomeration or in medium-sized cities. We see clearly the risk of erroneous inference when the individual-level variance is poorly specified.

Conclusion

We have now examined and described the main models that can be used in demography when discrete data are available. However, we may need to deal with other situations, which we shall outline briefly here.

The first possibility concerns panel data, which may involve discrete responses. As with the regression model analyzed in the previous chapter, we can study such data with multilevel methods. We refer the interested reader to the works by Diggle et al. (1994) and Singer and Willet (2003), which discuss these issues in great detail.

In other cases, we may want to model discrete and continuous characteristics simultaneously. One example is when we have data on (1) the proportions of smokers and non-smokers in a population and (2) the mean number of cigarettes smoked daily by smokers (Duncan, 1997). Again, we can estimate such models with multilevel methods (Goldstein, 2003).

Alternative estimation methods can also be used, such as Bayesian methods applicable to multilevel models. This is, in fact, a totally distinct branch of estimation in statistics, which would alone require a complete book-length description. We have chosen here not to elaborate on these methods, which would take us too far, and to discuss them more fully in another, future work.

CHAPTER VIII

MULTILEVEL EVENT-HISTORY ANALYSIS

We now arrive at the fullest demographic analysis, which uses the individual's entire life history to explain behaviors whose dates of occurrence vary with each individual. This analysis will, of course, rely on different types of segmentations, whose effect on the behaviors observed will be analyzed as well.

The first type of segmentation concerns repetitive events: a multilevel model will be able to treat each episode as a first aggregation level, the second being the individual. For instance, intervals between successive births for an individual woman will constitute the first level, and the set of women studied the second level. Likewise, we can examine a person's different migration stages or occupational stages, including periods of inactivity or unemployment. This will enable us to identify individual heterogeneity and thus solve the problem of unobserved heterogeneity, for which virtually no satisfactory solution exists when we observe only a single event per individual (see chapter III). Now, we have an infinite number of models incorporating different distributions of these unobserved phenomena, which adjust identically to the observed data. Only when we can observe different events of the same type (e.g., fertility, migrations, and successive unions) shall we be able to estimate this unobserved heterogeneity in each individual with certainty (Lillard, 1993; Delaunay, 2001).

The second type of segmentation is used when only a single event or a set of non-repetitive events (marriage and exit from agriculture, for example) is observed per individual, but when they occur in more aggregated units whose effect we want to illustrate. For intra-regional stays, the segmentation will be regional; for intra-municipal stays, we should apply a division into municipalities. However, that is not necessarily feasible, and we shall have to make do with a broader segmentation such as *départements* or regions (in France). Other, non-geographic segmentations may be employed. For example, a study of job durations calls for a segmentation into enterprises or employers.

Both types of segmentation can, of course, be used simultaneously. To return to the example of an individual woman's successive childbirths, we can add a third geographic segmentation—into regions—to the analysis. We can equally extend this type of model to cross-classifications or memberships in multiple structures.

We shall assume that readers are familiar with event-history models. Readers seeking fuller information on this approach before addressing multilevel models should refer to the manual on the subject published in French, English, and Spanish (Courgeau and Lelièvre, 1989, 1991, 2002), the volume in the "Méthodes et Savoirs" series (Lelièvre and Bringé, 1998) or, for more statistical aspects regarding these models, to the book by Andersen *et al*. (1993),

which draws on counting-process theory to offer a coherent examination of event-history models. We shall therefore discuss event-history models only briefly here, in order to concentrate on the issues raised by multilevel event-history analysis.

We begin with the existing sources and those that need to be established for the purpose of conducting a true multilevel analysis of event-history data; there follows a theoretical presentation of the models, which we shall apply to observed data.

I. - Existing data and sources to establish

So far, we have not addressed the issue of appropriate sources for multilevel analysis. The standard surveys may record either the individual's geographic location (confidentiality rules will generally exclude the use of exact addresses; in most cases, the *département* or region of residence will have to suffice), or the individual's presence in a given grouping at the time of the interview (family, school class, enterprise, etc.). If these criteria are met, we can conduct a period analysis of the population surveyed.

Now that we want to perform a multilevel event-history analysis, we need to be able to locate individuals throughout their lives. Various sources allow this in varying degrees, depending on their exhaustiveness and the information-gathering method.

The population register—when it exists and is properly maintained (Courgeau, 1988)—is a first-rate source for multilevel analysis. It captures all migrations by individuals throughout their lives and the main family events. When matched against other administrative sources or different censuses via a single national identification number (Poulain, 1996), it offers greater flexibility of use. One example is Belgium's centralized population register, which tracks the spatial location of its population and allows very fruitful multilevel analyses (Goldstein *et al.*, 2000). We have even been able to access the Norwegian population register, whose data we use in conjunction with the 1971 and 1981 censuses, kindly supplied by Statistics Norway. However, for reasons of confidentiality, we were only allowed to view the different lifelong regions of residence for generations born in 1947 and 1957. Although the data enabled us to perform highly interesting multilevel logit analyses of these generations (Baccaïni and Courgeau, 1996; Courgeau and Baccaïni, 1997, 1998; Courgeau, 2002, Courgeau, 2003; and the present volume), we shall not use the source for event-history analysis in this chapter.

Such registers are, of course, maintained in other countries as well, particularly those of Northern Europe (e.g., Netherlands, Sweden, Finland, Germany, and Denmark), Central/Eastern Europe (e.g., Poland, Czech Republic and Slovakia, Romania, and Russia), Japan, and China. In some of these countries, however, register-keeping is far from perfect, ruling out a true multilevel analysis.

In addition, the matching of register data and other sources is very costly and often impossible due to the existence of different identification numbers for each source. However, there are registers for land tenure, dwellings, employment, social security, taxes, defense, businesses, and trade (Trollegaard, 1995). In Denmark, for example, statisticians have been able to link the population register to workplace statistics and thus measure commuting patterns (Thygesen, 1983).

In many other countries that lack population registers, surveys are needed for lifelong tracking of individuals. Indeed, even in countries that maintain registers, such surveys are necessary to capture events not recorded by the central government, such as entry into

cohabitation and change of workplace. The survey sample must be large enough to identify an adequate number of areas or groupings of individuals. There are two types of surveys, depending on whether the data are collected over time (prospective survey) or at a particular moment (retrospective survey). We shall describe them here briefly, referring the reader to Courgeau and Lelièvre (1988, 1991, 2002) for more details.

The prospective survey follows the individual over time in order to obtain good-quality information on the dating of selected events and their memorization. Françoise Cribier's surveys of two cohorts of retired persons in France belong to this category (Cribier and Kych, 1999). The downside is that tracking a sample over a long period will cause losses of individuals who will have moved without leaving their new address or who will refuse to continue the survey. These "attrition" losses will not be random, but will form a subset of very specific characteristics (divorced persons, unstable individuals, etc.): eventually, the sample will thus cease to be representative of the population surveyed. Lastly—and most discouragingly—researchers who undertake such a survey know that they will be unable to use its results for a possibly very long time in order to effectively capture a "slice of life" large enough for multilevel event-history analysis. For instance, despite the fact that Françoise Cribier's investigation concerned a cohort of retired persons, it is not certain that the last respondent was dead 40 years after the survey. The researcher will accordingly prefer to use a retrospective survey that will supply directly usable results—albeit of lower quality, as we shall see.

The retrospective survey provides all the collected event histories immediately after its execution. Since it involves only one stage, there is no risk of losing individuals during the survey because of fatigue, death, and other causes. On the other hand, the quality of the information gathered retrospectively may be deficient: some events may have been forgotten, others misdated. Hence the need to determine the risk of information loss for a retrospective survey. After conducting the "Triple Biography" survey (family, migration, and occupation), we were able to perform a similar survey in Belgium—a country with population registers—to test the quality of retrospective information (Poulain *et al.*, 1991; Courgeau, 1999). Despite numerous misdatings (Auriat, 1996), many event-history analyses conducted on these four samples (each partner taken separately, both partners interviewed jointly, and the population register) produce highly convergent results that cannot be discerned from one another. We therefore conclude that dating errors act as background noise from which we can extract consistent information regardless of the source used. Memory is therefore reliable where analysis requires it to be.

Such surveys may also fail to collect all the information on the individual's successive locations. The "Young People and Careers" survey by INSEE in France captured family, occupational, and migration events for a sample of nearly 20,000 people. However, for cost reasons and because of the survey's excessive duration, while all the events were captured, their location in geographic space (for migrations) and in work space (for occupational changes) were not. Only some locations (around age 15, for example) and some enterprises where sample members worked (instantaneous data points 7 years apart) were recorded. We can thus see the major information loss suffered in this gathering process by comparison with a register population matched against an employment register. Nevertheless, as shown in this manual, the survey is a usable source.

To conduct a true multilevel event-history survey would require setting up an observation system representative of diversified and partly hierarchized social contexts (Loriaux, 1989). Unlike the situation in the biological sciences, human societies are

characterized by different aggregation levels not all of which are hierarchized—far from it. The complexity of human organizations, in which notions of *frontier* or *organization level* are much more blurred, make this task difficult. It would be useful to undertake research on the best way to optimize observation plans in order to take account of the main levels found in a human society: from individual to household, enterprise, religious community, and so on. As noted earlier, sources do exist on these aspects of social life, but the goal here is to consolidate them in a consistent manner so as to refocus these statistics on major social groups, which constitute society. This seems hard to achieve, but the stakes are high for understanding our societies.

II. - A two-level model for event-history analysis

Let us briefly recall the basic concepts of event-history analysis here. We follow the course over time of one or several processes, and we therefore know the individuals' lengths of stay in the initial state until they experience the event(s). Let us suppose that we restrict our study to the occurrence of one event, which we can naturally generalize to any number of events. From an initial instant assumed to be equal to zero, the event's date of occurrence, T, will be a positive random variable whose distribution we shall be able to examine. Let us suppose here that the variable is continuous, but it is equally possible to introduce discrete time.

A first useful notion is the survivor function: we define it as the probability that T is at least as great as a value t:

$$S(t) = P(T \geq t) \qquad 0 < t < \infty$$

(VIII.1)

This function is non-increasing, and left-continuous with $S(0) = 1$. As not all the individuals may experience the event, the function's limit when t tends toward infinity may be non-null: statisticians often resort to the mathematical artifice of a point at infinity to ensure that $\lim_{t \to \infty} S(t) = 0$.

A second notion is the instantaneous rate of failure at T = t conditional upon survival to time t, also called hazard function, defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < t + \Delta t \mid T \geq t)}{\Delta t} = -\frac{d \log S(t)}{dt}$$

(VIII.2)

A useful third notion is the cumulative or integrated hazard until date t, i.e., the integral of the conditional density:

$$H(t) = \int_{s=0}^{t} h(s)ds$$

(VIII.3)

It is important to realize that, while this notion may not seem fundamental if we confine our study to the occurrence of a single phenomenon, it becomes critical if we study the occurrence of competing processes simultaneously: when these risks are not independent, they allow a clear comparison of their effects, unlike survivor functions (Andersen et al. 1993).

To factor in the effect of various characteristics on the length of stay, we need to formalize this type of model more fully. We shall use proportional-hazard regression models, which usually fit the demographic data well. They assume that individual characteristics act multiplicatively on a hazard, which is identical for the entire population, throughout time. As a

result, the individual hazard functions are all proportional among themselves, whatever the duration studied. If $h_0(t)$ represents this underlying hazard, whose form can be either parametric (exponential, Gompertz, Weibull, Gamma, log-normal distribution, etc.) or fully non-parametric (semi-parametric Cox model), the hazard function for an individual possessing the explanatory characteristic x_1 will be of the form:

$$h(t;x_1) = h_0(t)\exp(a_1 x_1)$$

(VIII.4)

where a_1 is a parameter to be estimated. We see that when the individual does not possess the characteristic x_1, his or her hazard function is reduced to $h_0(t)$, and when he or she possesses the characteristic, the function is equal to $h_0(t)\exp(a_1)$. The relationship between the hazards of individuals who possess the characteristic and those who do not is equal to $\exp(a_1)$, interpreted as a relative risk,[12] i.e., a time-independent value. Accordingly, this model is called a proportional-hazards model. A convenient way to determine whether a model meets this description is to plot the log of the cumulative hazards for the different values of the characteristic as a function of time: the curves obtained should be parallel.

When a proportional-hazards model is not confirmed, we can naturally estimate other types of models. Accelerated failure time (AFT) models are often used, in which the characteristics exert a multiplicative effect on the length of stay itself (Bagdanovicius and Nikulin, 2002).

By introducing an aggregation level j in addition to the individual level i, and continuing to include an explanatory characteristic for this length of stay, now written x_{1ij}, we can write the simplest multilevel model as:

$$h(t_{ij};x_{1ij}) = h_0(t_{ij})\exp(a_1 x_{1ij} + u_{0j})$$

(VIII.5)

where the underlying probability will be multiplied by a term $\exp(u_{0j})$ dependent on individual i's region of residence j. This gives a proportional effect of the region on the hazard function as well as the cumulative hazard. Naturally, it will be possible to introduce more complex dependencies between the underlying probability and the region, or between the explanatory characteristic and the region. We shall thus be able to estimate a fuller model than model (5): it will incorporate a random variable at the explanatory-characteristic level, which we can write as:

$$h(t_{ij};x_{1ij}) = h_0(t_{ij})\exp(a_1 x_{1ij} + u_{1j}x_{1ij} + u_{0j})$$

(VIII.6)

where the underlying probability still depends on the region j via the term u_{0j}, just as the effect of the characteristic will also depend on this region via the term $u_{1j}x_{1ij}$.

To include non-proportional dependencies between the underlying probability and the region, it is often useful to model the probability, for example by using a polynomial exponential function with a sufficient degree to adjust it correctly. We can thus write the probability as:

$$h(t_{ij};x_{1ij}) = \exp[b_0 + v_{0j} + (b_1 + v_{1j})t_{ij} + (b_2 + v_{2j})t_{ij}^2 + \cdots + (a_1 + u_{1j})x_{1ij}]$$

[12] If the characteristic is dichotomous, it separates individuals into two sub-groups and we effectively measure an individual's relative risk compared with the base category. If not, we can say that when the value of x_1 increases by one unit, an individual's hazard function is multiplied by $\exp(a_1)$.

where the parameters $b_0, b_1, b_2 \cdots, a_1$ are to be estimated, along with the variances and covariances of the random variables. It is generally preferable to work on values centered around the mean durations in order to avoid excessively high values for successive powers and hence excessively low values for the estimated parameters, but that would not change the conclusions of the analysis in any way.

We can easily extend this formalization to event-history models with competing risks, where the event studied can occur in different ways or for different reasons: halt in the use of a contraceptive method for various reasons (Steele et al., 1996), departure from parental home due to marriage, cohabitation or other reasons (Courgeau, 2000), mortality by cause of death, etc. It can also apply to the study of multivariate phenomena, for which individuals may go through different stages, not necessarily ordered. Two examples are the study of links between nuptiality and departure from agriculture using a bivariate model (Courgeau and Lelièvre, 1985), and the study of twin deaths (Hougaard et al. 1992). This set of models is analyzed in detail in the work by Hougaard (2000).

Let us now examine in greater detail some of the conditions that ensure the models' validity.

One distinctive feature of the models is that some individuals have not experienced the event studied during the observation period or may never even experience it at all. These constitute exits from observation by individuals who have not yet experienced the event. Event-history models are capable of taking account of such right-censored intervals, when the end-of-observation mechanism is random or non-informative about the event studied. By contrast, it is far more difficult to handle left-censored intervals when we do not know the start of the observation. This can occur in the study of diseases such as AIDS, where it is difficult to date the infection time exactly, our only available datum being the instant of detection of the disease by a physician.

A second specificity of these models is their assumption that each duration observed corresponds to a single event. When this is not so, we can always use counting-process theory to aggregate a given number of processes in order to obtain a multiple-jumps process. This will no longer be a counting process, but we can still estimate its cumulative probabilities and their variance.

A third feature is that individuals have no reason to remain in the same area or category throughout the study, unless we select migrations between areas or categories as the principal phenomenon to be analyzed. While various solutions to this problem exist, we cannot address it properly without an underlying theory that, for the moment, is difficult to test.

We may begin by assuming that individual behavior is very slow to adjust to that of the area or category of destination, and we may suppose that migrants keep the behavior of the area of origin. If so, we can conduct the analysis without taking these migrations into account. On the contrary, we can assume that the adjustment of individual behavior to that of the area or category of destination is immediate. If so, the analysis will need to transfer the individual from a population at risk to another immediately after the migration. This hypothesis of instant adjustment to the behavior of the area of destination may be confirmed when living conditions and costs in the locality enable the individual to fulfill certain desires that could not be satisfied in the place of origin. For instance, the change in fertile behavior of women initially residing in a highly urbanized area who have migrated to a lower-density area may be explained by living conditions (the women can stop working) and dwelling size and cost, made easier after migration (Courgeau, 1987).

A third hypothesis should, in fact, be more widely followed. While an individual's change of behavior just after migration generally seems unlikely, there are grounds for assuming that a longer-term adjustment to the behaviors of the area or group of destination

should occur. We would then have to suppose that individual behavior shifts from one type to another over a period to be determined, and at a pace that is not necessarily identical throughout the period. Such a determination calls for special surveys, which have barely been conducted yet. For the time being, therefore, this approach seems hard to apply.

III. - Estimating the model's parameters

Many publications have been devoted to the estimation of multilevel event-history models (Clayton and Cuzick, 1985; Hougaard, 1986; Klein, 1992; Sastry, 1997 ; Sargent, 1998; Vaida F., Xu R., 2000; Ma et al. 2003). Here we examine the very simple case of model (5), where the random variable to be estimated, $\exp(u_{0j})$, is a regional term with a multiplicative effect on the hazard function. We know that for such a model we cannot directly maximize conventional likelihood, as the underlying probability $h_0(t)$ is not specified. Instead, we use a partial likelihood of the type proposed by Cox (1972).

Let us begin by assuming that only a single event occurs at each date t_{ij} throughout the observation period. Conditionally upon the population at risk and an occurrence on date t_{ij}, the probability that the individual i present in area j will experience the event is equal to:

$$\frac{h_0(t_{ij})\exp(a_1 x_{1ij} + u_{0j})}{\sum_{l,k \in R_{ij}} h_0(t_{lk})\exp(a_1 x_{1lk} + u_{0k})} \qquad \text{(VIII.7)}$$

where R_{ij} stands for the set of labels of individuals at risk in $t_{ij} - 0$. We easily see that this expression reduces to:

$$\frac{\exp(a_1 x_{1ij} + u_{0j})}{\sum_{l,k \in R_{ij}} \exp(a_1 x_{1lk} + u_{0k})} \qquad \text{(VIII.8)}$$

We can thus ignore the intervals in which no event occurs and for which our only information is the exit of individuals from observation. We assume that these exits occur immediately after the event. The information supplied by a precise date of exit from observation is assumed to be negligible. The partial likelihood is formed by taking the product of the preceding expressions on all dates:

$$PL = \prod_{t_{ij}} \frac{\exp(a_1 x_{1ij} + u_{0j})}{\sum_{l,k \in R_{ij}} \exp(a_1 x_{1lk} + u_{0k})} \qquad \text{(VIII.9)}$$

We see that, conditionally to the values of x_{1ij}, each of the VIII.8 probabilities is independent of the others.

The hypothesis that only a single event occurs at each observation date is too restrictive: in most of the samples observed, several event very often occur on the same date. If we note d_{ij} the number of individuals included in the set D_{ij} for which the date of occurrence

is t_{ij}, we have several approximations of partial likelihood (Cox, 1972; Peto and Peto, 1972; Kalbfleish and Prentice, 1973). Here is the one given by Peto and Peto (1972):

$$PL = \prod_{t_{ij}} \frac{\exp(a_1 s_1 + u_0)}{\left[\sum_{l,k \in R_{ij}} \exp(a_1 x_{1lk} + u_{0k}) \right]^{d_{ij}}} \qquad \text{(VIII.10)}$$

where

$s_1 = \sum_{i,j \in D_{ij}} x_{1ij}$ and $u_0 = \sum_{i,j \in D_{ij}} u_{0j}$.

Note that this approximation assumes a low number of simultaneous occurrences. Kalbfleisch and Prentice's estimation is obtained by an accurate processing of these multiple occurrences but is difficult to implement with existing software.

As we can see, this partial likelihood, which is the product of the contributions of each event examined, differs from standard likelihood, which is the product of each individual's contributions to the sample. This decomposition entails an estimation of the parameters of a Poisson multilevel model of the type described in chapter VII. If we define, at each moment l when an event occurs, a response of the form:

$$y_{hij(l)} = \begin{cases} 1 & \textit{if } h \textit{ is the event observed} \\ 0 & \textit{if not} \end{cases}$$

where h indexes the members of the population at risk just before the event, we obtain an artificial Poisson model, which allows us to estimate the parameters of a three-level model. The first level corresponds to the event, the second to the individual, and the third to the region. This model, which introduces the mathematical mean of the response, can be written:

$$E(y_{hij(l)}) = \pi_{hij(l)} = \exp(\alpha_l + a_1 x_{1ij} + u_{0j}) \qquad \text{(VIII.11)}$$

where $\pi_{hij(l)}$ is the observed ratio between the number of events experienced and the population at risk just before the event considered, $N_{ij(l)}$, and α_l a term called "blocking factor" for the term $\log[h_0(l)]$. We can naturally estimate one term per observation duration, but it is often preferable to use a low-order polynomial or a spline function to model this blocking factor.

For an individual situated in the jth region at the instant t, whose characteristic x_{1ij} is equal to zero, the cumulative hazard is equal to:

$$\hat{H}_j(t) = \sum_{l \le t} \exp(\hat{\alpha}_l + \hat{u}_{0j}) \qquad \text{(VIII.12)}$$

and the estimator for the individual possessing the characteristic x_{1ij} will have to be multiplied by the term $\exp(a_1 x_{1ij})$.

Alternatively, we could have treated these responses with a multinomial model yielding the same estimates as the Poisson model: for more details on these estimations, see McCullagh and Nelder (1989).

We shall not go further in the estimation of the parameters of the multilevel event-history model. Other options include estimating discrete-time models, accelerated-life models, and so on. Readers seeking more details on these estimations should refer to the articles quoted at the start of this chapter.

We shall now examine the problems involved in an estimation on data actually observed by the "Young People and Careers" survey.

Example no. 10: Departure from parental home, seen by "Young People and Careers" survey, in France

We take the results of an article on departures from parental homes observed in France as a whole (Courgeau, 2000). We now introduce a segmentation into départements identical to the one used in example no. 8, which groups together certain contiguous départements whose observed population was too small. Here, we define the first independent dwelling as the one where the individual has become, for the first time, a tenant, owner-occupier, or person housed by his or her employer. We confine our analysis to women's behavior, setting aside men. Also, we do not cover all characteristics used in the article, for our purpose here is to give a simple presentation of the estimations, describing the problems that we may encounter in different cases.

The characteristics included here are the following: (1) generation of the individual, born between 1952 and 1973, centered on 1962 and divided by 10; (2) number of siblings, when there are fewer than four, and the number is set to four for every additional sibling (we chose this segmentation as the most efficient for coding the characteristic, and it may be regarded as an approximately continuous variable); (3) the fact that the respondent's father and/or mother are foreigners (binary variable); (4) the fact that the mother is dead, regarded as a time-dependent variable, equal to unity from the year of death onward; (5) the fact that the father was a farmer, and (5a) the proportion of farmer fathers in the respondent's département of residence, which enables us to include a characteristic and its aggregate equivalent simultaneously; (6) the respondent's previous work spells or unemployment spells, which also constitute time-dependent characteristics.

Let us begin with an overall analysis of these first departures, before classifying them into three types (due to marriage, start of cohabitation or other reasons). We shall use a degree-five polynomial law to represent the duration-of-stay effect, which is well suited to the data.

Table VIII.1 gives the results of this first analysis in three stages, which we shall present in greater detail.

TABLE VIII.1. - TOTAL DEPARTURES FROM PARENTAL HOMES AS A FUNCTION OF DIFFERENT INDIVIDUAL CHARACTERISTICS

Fixed effects:	Model 1	Model 2	Model 3
a_0 (constant)	-1.529 (0.030)	-1.671 (0.041)	-2.148 (0.046)
a_1 (generation)	-0.191 (0.022)	-0.167 (0.021)	-0.046 (0.023)
a_2 (generation^2)	-0.277 (0.036)	-0.269 (0.036)	-0.248 (0.037)
a_3 (number of siblings)	-	0.058 (0.009)	0.043 (0.009)
a_4 (foreign father or mother)	-	-0.275 (0.036)	-0.229 (0.036)
a_5 (mother's death)	-	0.105 (0.062)	0.131 (0.062)
a_6 (farmer father)	-	-0.060 (0.040)	-0.100 (0.040)
a_7 (% of farmer fathers in département)	-	0.639 (0.192)	0.752 (0.182)
a_8 (previously worked)	-	-	0.629 (0.027)
a_9 (previously unemployed)	-	-	-0.292 (0.057)
Random variable, département level: σ^2_{u1} (generation)	0.006 (0.005)	0.005 (0.005)	0.007 (0.006)
Random variable, period level: σ^2_{e0}	1	1	1

These different models give us, in addition to the parameters listed in this table, duration-of-stay effect (duration t is centered here on a 12-year period ($\theta = t - 12$), to avoid excessively large terms at higher powers). We set aside the constant-term effect already supplied in table VIII.1, narrowing our focus to the time effect, modeled by the following polynomial function:

$$\log[h_0(\theta)] = (\nu_{1j} - 0.07269)\theta - 0.002137\theta^2 - 0.0004224\theta^3 - 0.0001165\theta^4 + 0.0000112\theta^5$$

We have had to go up to the fifth degree to ensure that the following term had no significant effect. The duration of stay itself influences the random variables at département level quite significantly: the term ν_{1j} indicates that probabilities vary as a function of time in different ways from one département to another. This enables us to estimate the cumulative hazard or survivor functions for the departure from parental homes for each département. The only reason why we have examined both cumulative hazard and survivor functions is for comparison purposes. As it is impossible to plot all their curves in this chapter, we shall present them in a more summary fashion.

Figure VIII.1 plots the survivor functions and figure VIII.2 the cumulative hazard, for the generation born in 1962, for three types of départements: the first is represented by Alpes-Maritimes (similar to Paris, Bouches-du-Rhône, Haute-Garonne, Gironde, Seine-Saint-Denis,

and Val-d'Oise), with a very rapid departure from the parental home; the second is represented by Calvados (similar to Aube, Ardennes, Dordogne, Manche, Finistère, Maine-et-Loire, and Sarthe), with a very slow departure from the parental home; the third is represented by Côte-d'Or (similar to Hérault, Indre, and Vienne), with an intermediate profile. However, when the duration of stay increases, all these curves converge to produce a virtually identical, near-unity intensity for all the départements, as shown by curve VIII.1.

For a closer examination, we have plotted on map VIII.1 the values of the duration-of-stay random variable for all départements. A high positive value of the random variable denotes an early departure from the département, and a strongly negative value indicates a late departure. We have specified five classes: under –0.016 (very late departure), between –0.016 and –0.002 (late departure), between –0.002 and +0.010 (mid-length departure), between +0.010 and +0.021 (early departure), and over +0.021 (very early departure). The map confirms the earlier results by locating them more clearly on French territory. Most départements with the earliest departures (darkest shades on the map) are in southern France, except for Paris, Seine-Saint-Denis, and Val-d'Oise: they include the départements of the Alps, the Pyrenees, and the South-West. Most of them are also very strongly urbanized. The départements with the latest departures (lightest shades on the map) comprise the départements of Normandy and the Loire valley, Dordogne, Ardennes, and Aube.

Model 1 also shows a major fixed effect of generation and a lesser effect on département random variables. However, we have kept these elements here, for the random-variables effect is quite significant in some départements. The fixed effects produce a quadratic function showing an initial increase in the earliness of departures, during the first quarter of the period studied, followed by a steady decrease (figure VIII.3). This corresponds to chart I of the article cited earlier (Courgeau, 2000), which shows a decrease in the median age at departure from the parental home in the generations born between 1952 and 1960, followed by a continuous rise in the departure age for generations born after 1960. For a more precise view of the situation when performing a multilevel analysis, we have plotted these département effects on map VIII.2. A high positive value of the random variable denotes a late change in behaviors (lightest shades on the map), and a strongly negative value indicates a rapid change (darkest shades on the map). We have specified five classes: under –0.06 (very rapid change), between –0.06 and –0.03 (rapid change), between –0.03 and +0.01 (medium-speed change), between +0.01 and +0.05 (slow change), and over +0.10 (very slow change). This map is practically the opposite of the previous one, with a –0.85 correlation between the two random variables: the earlier the individual leaves the parental home, the sooner the change of behavior.

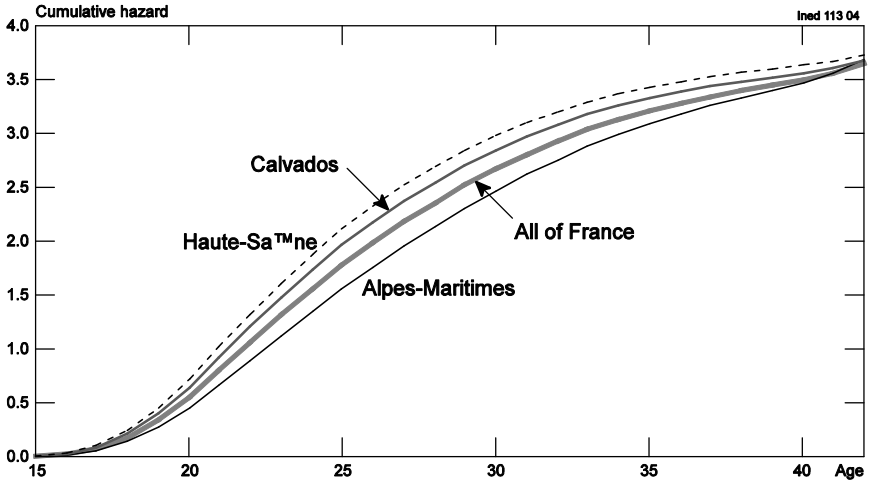


Figure VIII.1. - Survivor function for stay at parental home from age 16 on, in all of France and in three types of départements

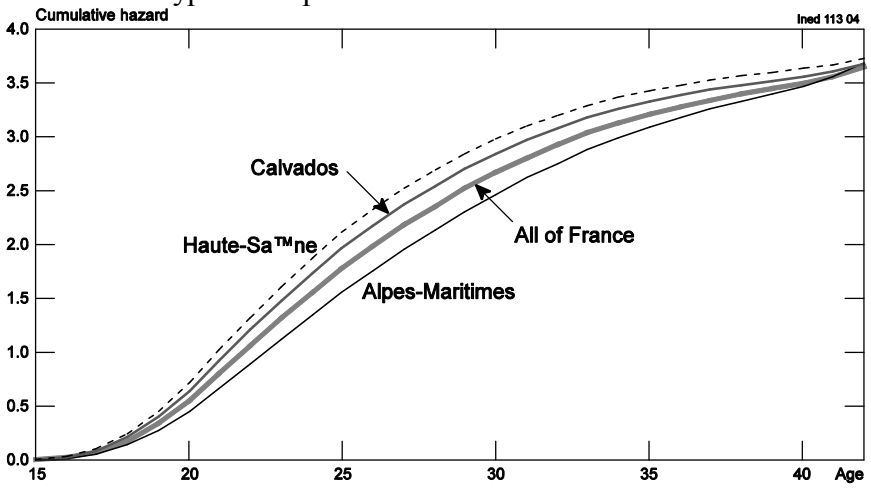


Figure VIII.2. - Cumulative hazard of departure from parental home from age 16 on, in all of France and in three types of départements

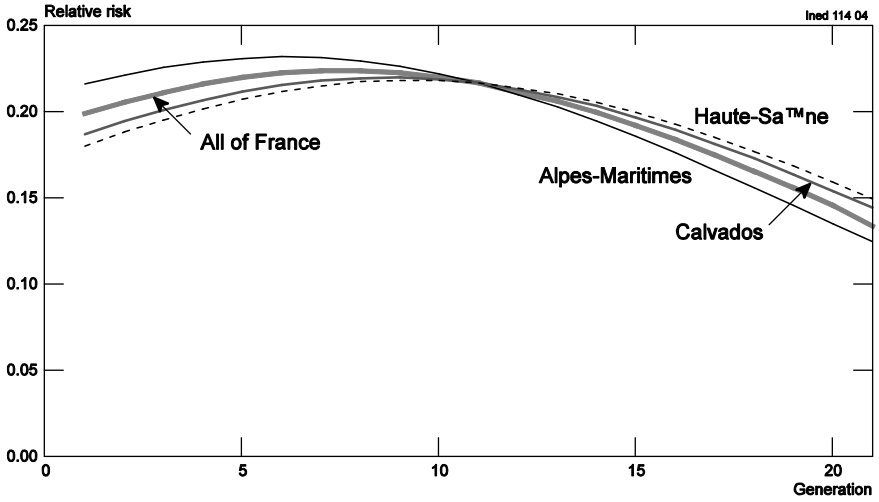
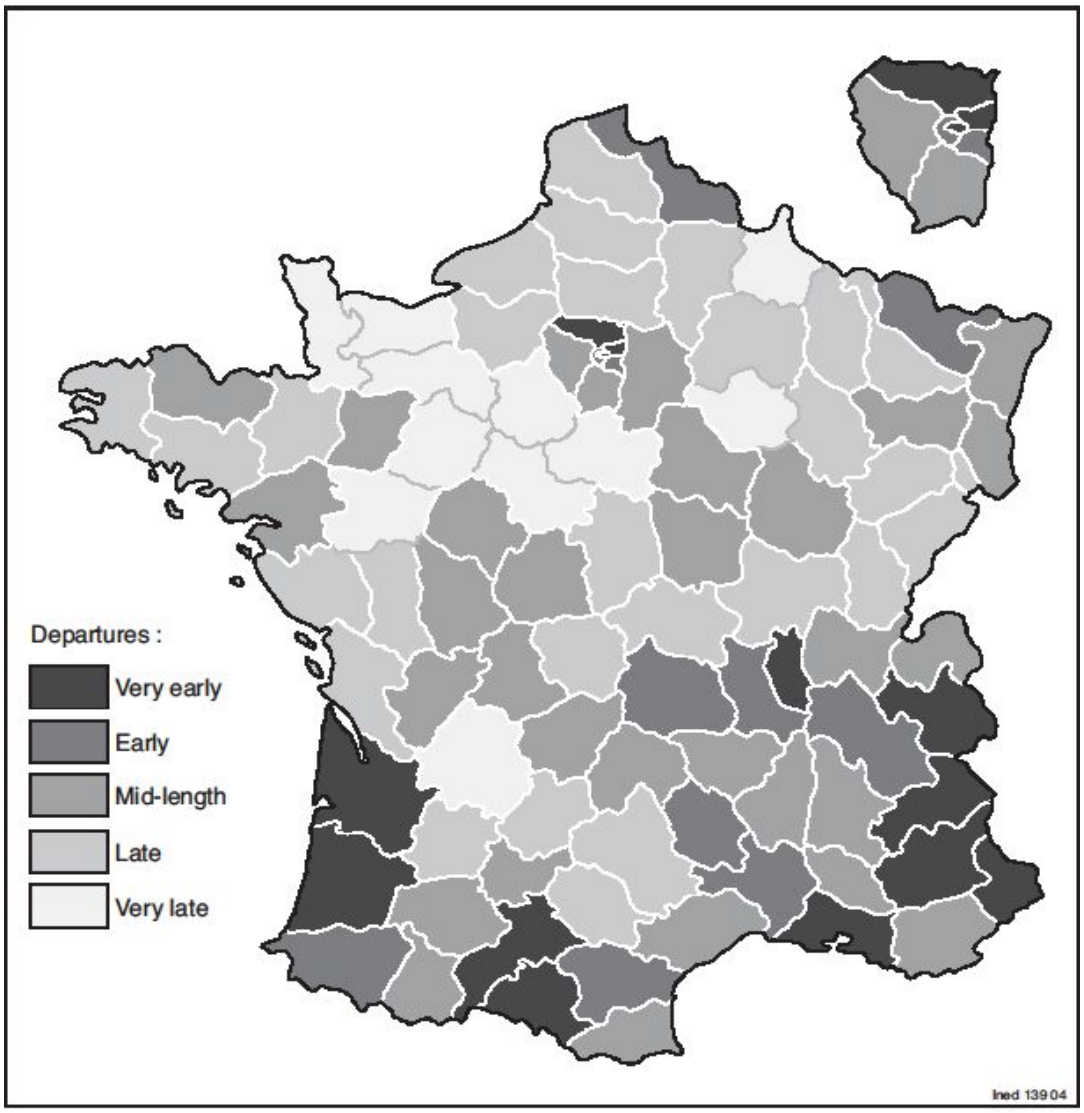
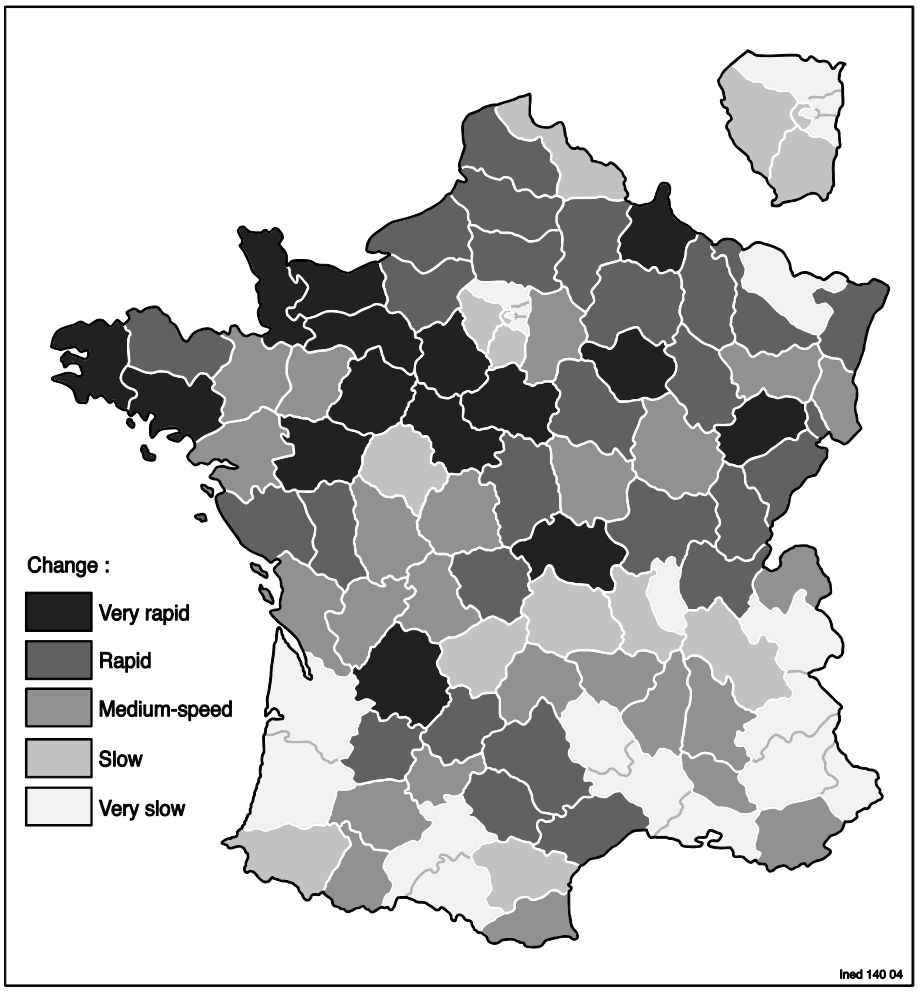


Figure VIII.3. - Relative risks as a function of generation in all of France and three types of départements



Map VIII.1. - Effects of duration of stay on departure from parental home



Map VIII.2. - Effects of generation on departure from parental home

Model 2 incorporates some family characteristics. The number of siblings has a fully significant effect: the larger the family, the earlier the children leave. Given the coding of this characteristic, we see that the probability of departure increases by about 6% per additional child to stabilize at 26% for four or more children. Having a foreign father and/or mother slows departures strongly. By contrast, the mother's death (when it occurs) increases them. In fact, we can use it as a time-dependent variable here: as the estimates are calculated by length of stay, we can simply introduce this characteristic once it occurs. While having a farmer father slows the children's departures, living in a département with a high percentage of farmers will quicken them. This constitutes a strong contextual effect running counter to the individual effect, as in our earlier example of migrations by Norwegian farmers (see chapter IV).

Model 3 includes some occupational characteristics of the woman herself. Again, these are time-dependent characteristics, for they involve events that can hasten or slow the departure from the parental home. Finding full-time work will allow a rapid departure. Conversely, a long unemployment spell will encourage a woman to remain with her parents.

The inclusion of these characteristics has no effect on the random variables at département level corresponding to the generation effect. On the contrary, they strongly reduce the variables linked to the duration of stay: by half for model 2 compared with model 1, and by nearly half as well for model 3 compared with model 2. As a result, the curves plotting the survivor function and cumulative hazard by département will converge after the

introduction of the characteristics examined here, although they will remain significantly different.

We now classify departures by cause: cohabitation, marriage or other reasons, largely occupational. We regard these as competing risks worth comparing. To do this, we cannot continue using the survivor functions, as they do not provide clear information for comparison purposes. The only alternative is to use hazards or—better yet—cumulative hazards (Andersen et al. 1991).

Let us first see the generation effects reported in table VIII.2

TABLE VIII.2. - EFFECTS OF GENERATION ON DEPARTURES DUE TO COHABITATION, MARRIAGE OR OTHER REASONS

Fixed effects:	Departures due to cohabitation	Departures due to marriage	Departures due to other reasons
a_0 (constant)	-2.690 (0.063)	-2.980 (0.073)	-2.208 (0.057)
a_1 (generation)	0.904 (0.104)	-1.308 (0.050)	-0.131 (0.076)
a_2 (generation^2)	-1.027 (0.084)	-0.663 (0.089)	-0.071 (0.055)
a_3 (generation^3)	-0.499 (0.149)	-	0.271 (0.104)
Random variables, département level:			
σ_{u0}^2 (constant)	0.032 (0.013)	0.105 (0.031)	0.071 (0.024)
σ_{u01} (covariance)	-0.009 (0.017)	-	-0.018 (0.014)
σ_{u1}^2 (generation)	0.099 (0.042)	-	0.017 (0.013)
σ_{u2}^2 (generation^2)	-	0.139 (0.073)	-
σ_{u02} (covariance constant x generation^2)	-	-0.112 (0.043)	-
Random variable, period level:			
σ_{e0}^2	1	1	1

Once again, we have not listed the effect of the duration of stay on the hazard A brief description of the effect follows.

For departures due to cohabitation, the duration of stay influences only the fixed effects, again up to the fifth degree:

$$\log[h_{0c}(\theta)] = -0.08432\theta - 0.0000715\theta^4 + 0.00001732\theta^5$$

For departures due to marriage, we can write:

$$\log[h_{0m}(\theta) = -0.1801\theta - 0.01117\theta^2 + 0.0000214\theta^5$$

and we see that the duration of stay has no effect on the random terms. For departures due to other reasons, we can write:

$$\log[h_{0a}(\theta)] = (v_{2j} - 0.07059)\theta^2 - 0.001209\theta^3 - 0.00001348\theta^4 + 0.00001348\theta^5$$

There is now a random term for the second-degree term. We begin by comparing the cumulative hazards obtained for the three types of departures; for the third type, we set the random variable to zero. Figure VIII.4 plots the comparison.

It reveals a different timing for the three types of events. Until age 25, departures due to cohabitation and marriage display similar timings; at older ages departures due to cohabitation are by far the largest category. Departures for other reasons continue to outweigh

the two other types, varying in an almost linear pattern with age, and slowing from age 27 up, whereas departures due to cohabitation accelerate after that age.

Let us now take a closer look at the patterns in selected départements that effectively cover the range of situations observed in most others (Alpes-Maritimes, Calvados, and Haute-Saône compared with all of France) before presenting more summary maps. We compare the cumulative hazards for the three types of departures by cause: cohabitation, marriage, and other reasons (figures VIII.5, VIII.6, and VIII.7).

For all three types of departures, Alpes-Maritimes continues to rank in the lowest position by comparison with all of France, as was already the case in figure VIII.2, which did not distinguish between reasons for departure. Haute-Saône is in the top position for departures due to cohabitation and marriage, but lies in mid-range for other departures, a category where Calvados is by far the leader.

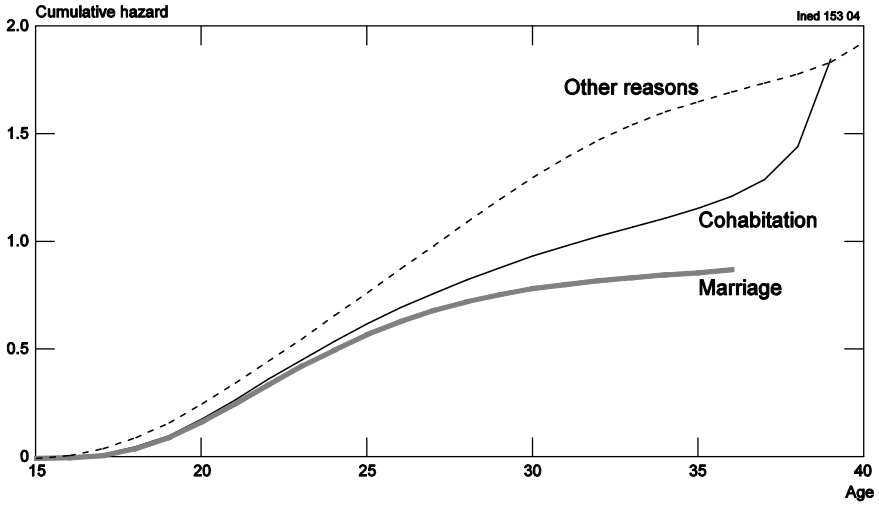


Figure VIII.4. - Cumulative hazard of departure from parental home due to cohabitation, marriage or other reasons

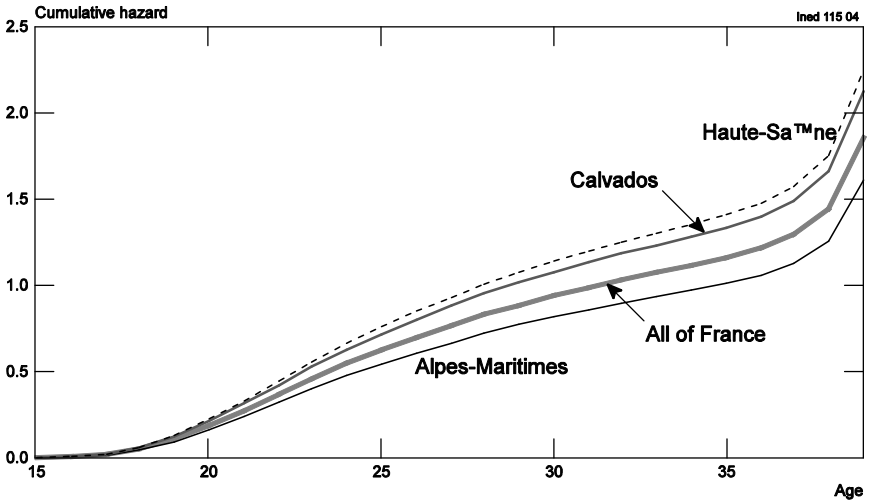


Figure VIII.5. - Cumulative hazard of departure from parental home due to cohabitation in all of France and three départements

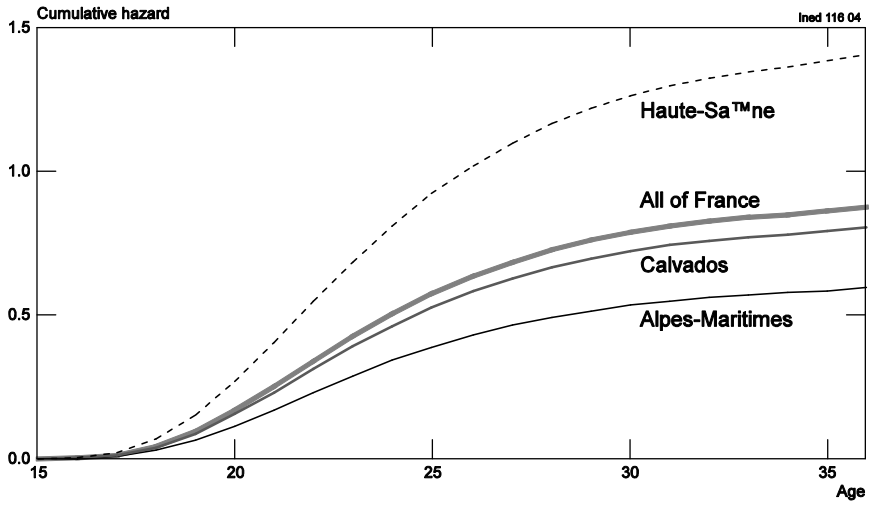


Figure VIII.6. - Cumulative hazard of departure from parental home due to marriage in all of France and three départements

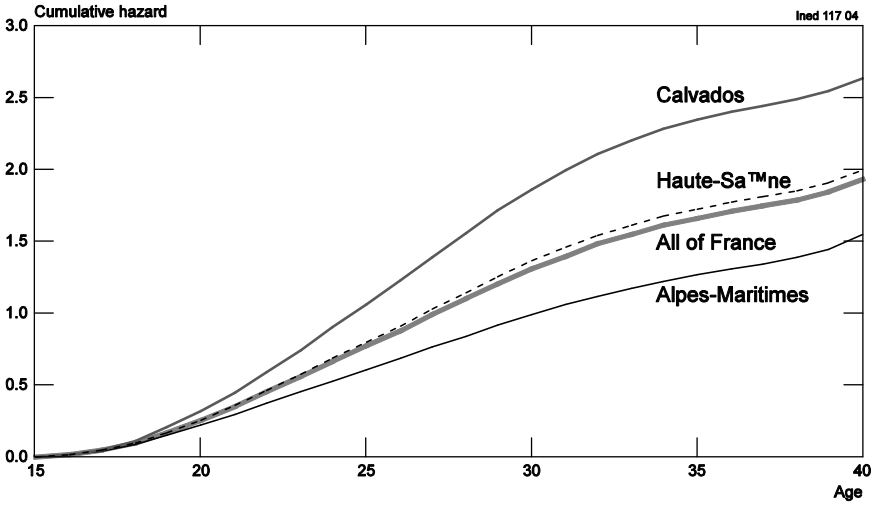


Figure VIII.7. - Cumulative hazard of departure from parental home due to other reasons in all of France and three départements

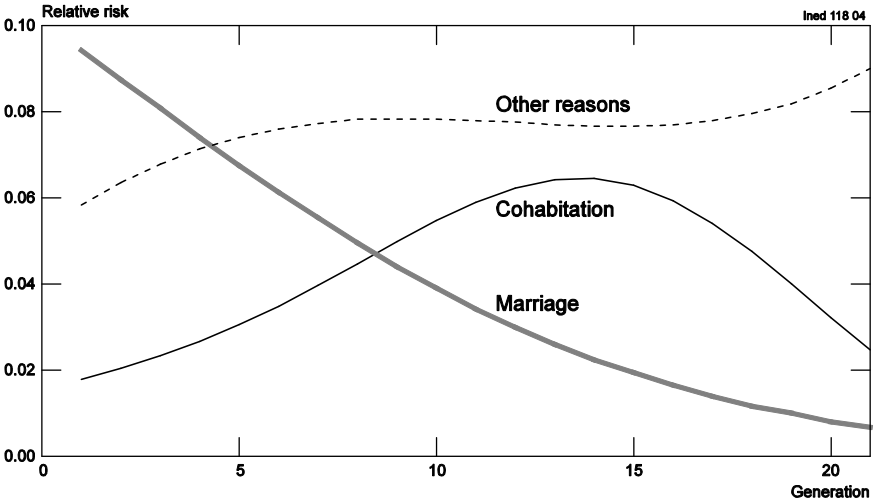


Figure VIII.8. - Relative risk of departure from parental home by reason in Alpes-Maritimes

In addition to these age effects, major differences appear between generations. By estimating relative risks, we can compare departures due to cohabitation, marriage, and other reasons. We calculate the risks with the fixed effects given in table VIII.2, to which we add, on a case-by-case basis, the random effects u_{0j}, u_{1j} and u_{2j} estimated for each département. They supply the constant terms by which we must multiply $h_{0c}(\theta), h_{0m}(\theta)$ and $h_{0a}(\theta)$ to obtain the probabilities for each type of departure and département of origin. Before providing an overview, let us examine three départements to see the ranking of the three types of departures.

Figure VIII.8 plots the relative risks for Alpes-Maritimes. While departures due to marriage predominate in the first generations, they are already strongly decreasing from the start of the observation. They are outranked by departures for other reasons starting with the generation born in 1955, and by departures due to cohabitation starting with the generation born in 1959. Departures due to other reasons remain stationary at a high level until the generations born in 1969, then rise. By contrast, departures due to cohabitation, after rising sharply until the generations born in 1964, later decline just as steeply.

Figure VIII.9 plots the relative risks for Calvados. Departures due to marriage and other reasons are at an identical high level for the generation born in 1952. But their curves then diverge, with departures due to marriage decreasing and departures due to other reasons increasing in later generations. The former peak in the generations born in 1960, while the latter decline continuously until the last generations observed. The pattern for departures due to cohabitation is identical to that of the Alpes-Maritimes.

Figure VIII.10 plots the relative risks for Haute-Saône. Departures due to marriage and other reasons start by increasing at identical rates until the generations born in 1957. Only later do departures due to marriage begin an uninterrupted decline, whereas departures for other reasons fluctuate at a high level. Departures due to cohabitation follow a similar path to the one observed in the previous two départements.

We can illustrate the spatial detail of these different types of departures with a series of maps.

Regarding departures due to cohabitation, figures VIII.8-10 indicate that they can be characterized by two maps: map VIII.1 gives the initial state of départements for the generation born in 1952, the first to register this type of departure from parental homes; map VIII.2 shows the peak number of departures due to cohabitation, before their renewed decline.

A very high value for the initial state indicates that the département had already recorded a sizable number of departures due to cohabitation in 1952; a very low value denotes near-zero departures of that type in 1952. We have divided departures into five classes: under 0.012 (very few departures), 0.012-0.015 (few departures), 0.015-0.018 (mid-range number of departures), 0.018-0.021 (large number of departures) and over 0.021 (very large number of departures). Map VIII.3 identifies several locations where departures due to cohabitation began early: the départements of the ring of suburbs around Paris, excluding Paris proper; the départements of the Alps, including the city of Lyon, the départements of South-West France, Côtes-du-Nord, and the départements of North-East France.

For the random variable linked to the peak number of departures due to cohabitation we have also defined five classes: under 0.068 (very low peak), 0.068-0.077 (low peak), 0.077-0.087 (medium-size peak), 0.087-0.095 (high peak), and over 0.095 (very high peak).

Map VIII.4 shows the size of the peak number of departures due to cohabitation. A large area extending westward (Normandy) and southward (Nantes, Le Mans, and Orléans) from the center of the Paris Basin registers the highest peak values. The lowest are in Paris, the Pyrenees, and the Alpes-Maritimes département.

For departures due to marriage, the peak is clearly visible in a sizable proportion of départements, while others had already reached their peaks prior to the start of the observation for generations born before 1952. We have therefore had to use an indirect measure provided by the parameters of the model used: the random variable for the square of the generation. A high positive value for this random variable indicates a very early start to the decline; conversely, a high negative value denotes a very late start. We have divided the values into five classes: under –0.3 (very late departure), between –0.3 and –0.1 (late departure), between –0.1 and +0.1 (mid-range departure), between +0.1 and +0.3 (early departure), and over 0.3 (very early departure). Map VIII.5 shows a decline in very early departures due to marriage in the départements of the Paris region, the Alps, and South-West France. By contrast, the decline occurs very late in the départements of the Vendée and Poitou-Charente regions, Northern France, and the Massif Central.

For departures due to other reasons, the preceding figures and the examination of the full set of results shows us that we can characterize them by the first peak reached for generations born between 1957 and 1965. This maximum value, ranging between 0.07 and 0.18, leads us to distinguish five classes: very low (under 0.09), low (0.09-0.11), medium (0.11-0.13), strong (0.13-0.15), and very strong (over 0.15). Map VIII.6 shows very low values in Normandy, Picardy, the Rhine Valley, the Mediterranean rim, and the Rhône Valley. The values are very high in certain départements of Brittany, Vendée, Poitou-Charente, and some départements of Central-Eastern France.

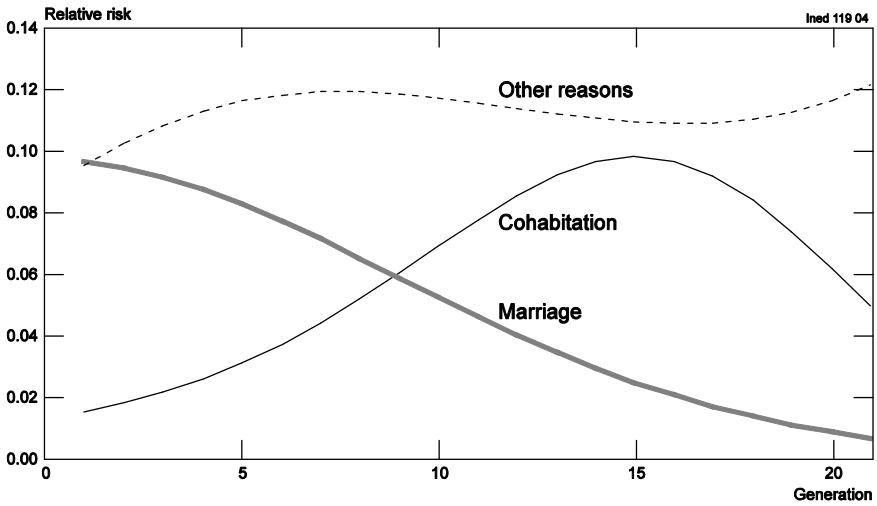


Figure VIII.9. - Relative risks of departure from parental home by reason for Calvados

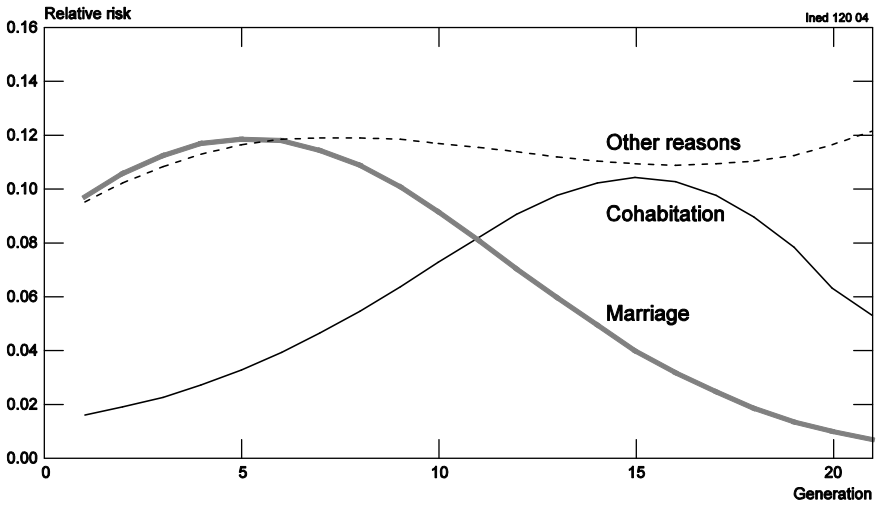
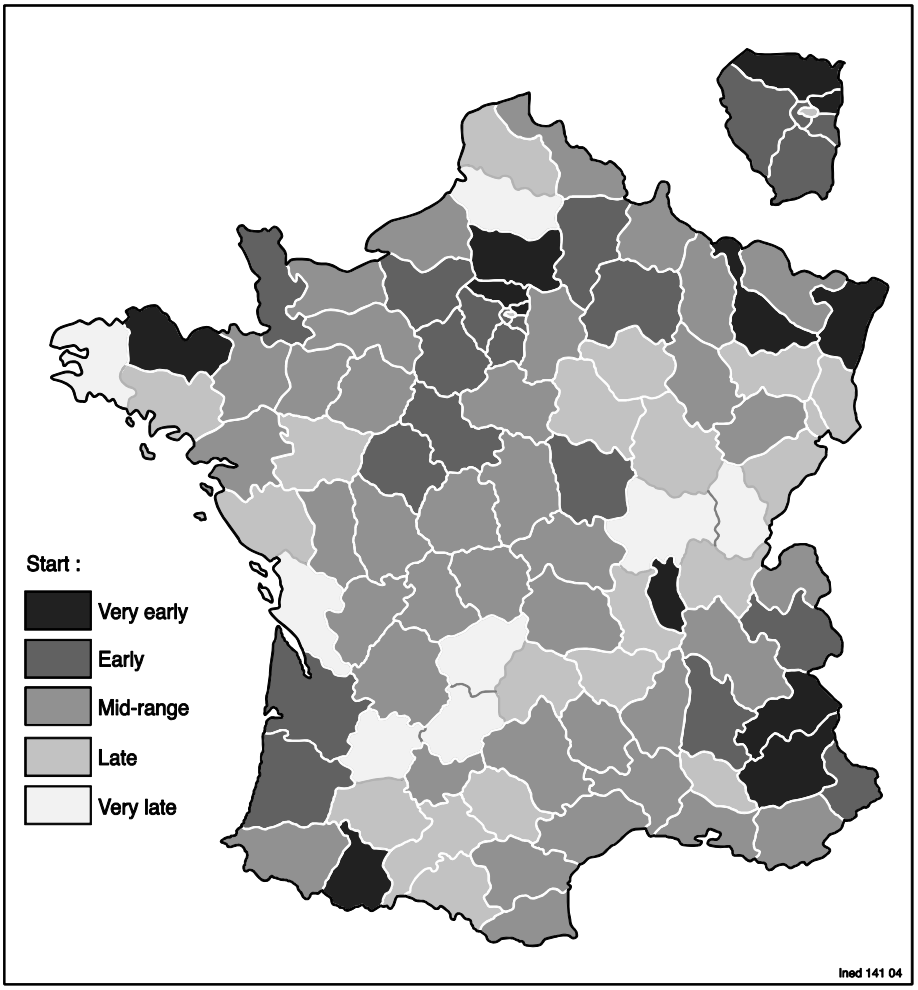
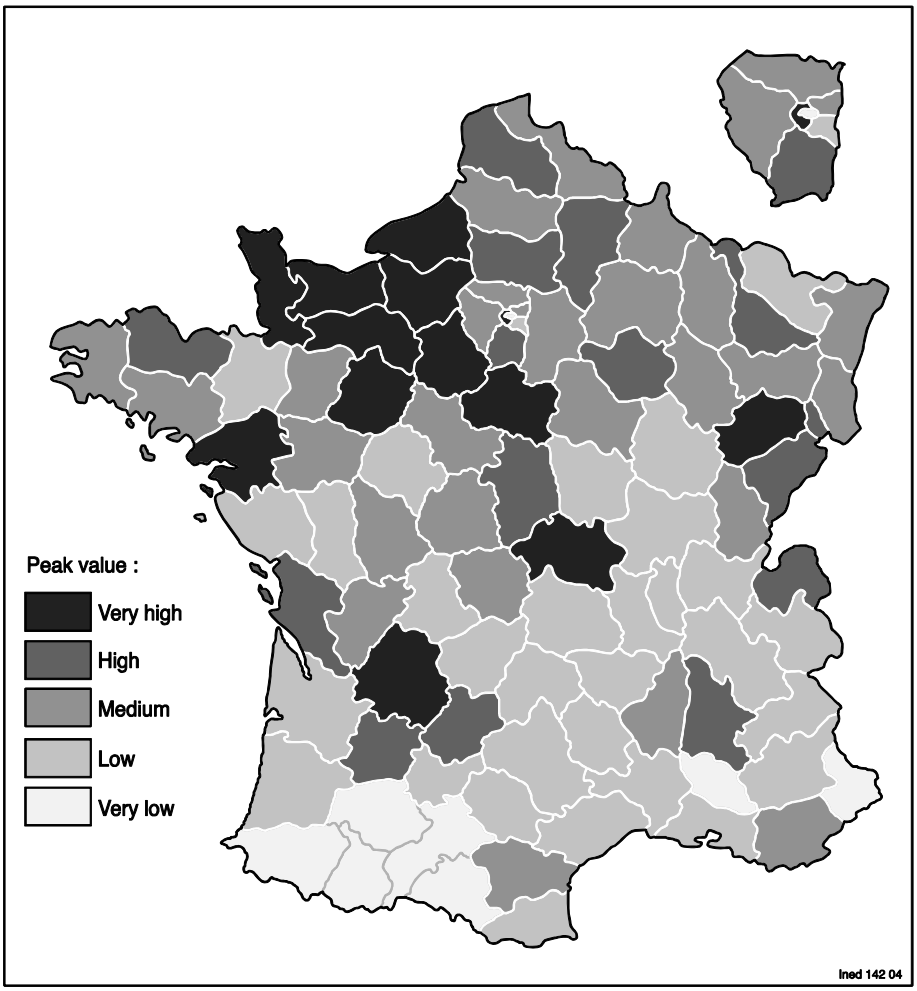


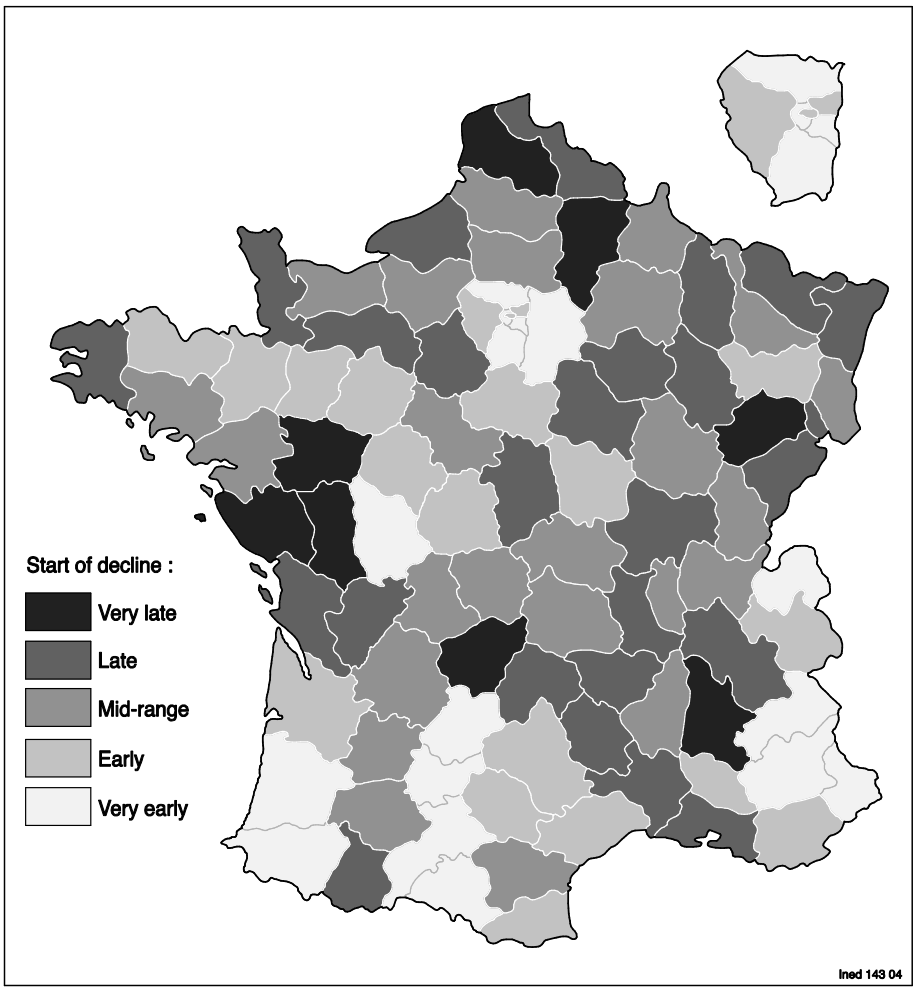
Figure VIII.10. - Relative risks of departure from parental home by reason for Haute-Saône



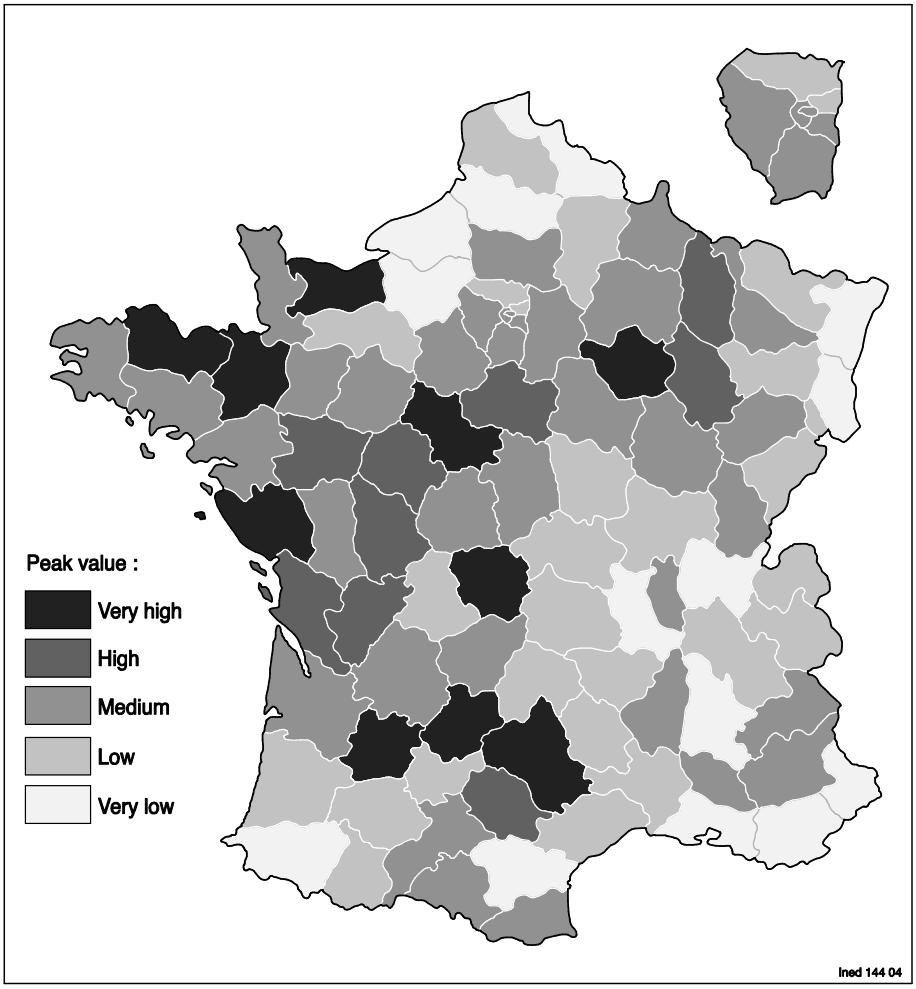
Map VIII.3. - Start of departures due to cohabitation



Map VIII.4. - Peak value of departures due to cohabitation



Map VIII.5. - Start of decline in departures due to marriage



Map VIII.6. - Peak intensity of departures for other reasons

Let us now simultaneously incorporate all the characteristics examined here. Table VIII.3 reports the results for the three types of departures.

TABLE VIII.3. - EFFECTS OF TOTAL CHARACTERISTICS ON DEPARTURES DUE TO COHABITATION, MARRIAGE OR OTHER REASONS

Fixed effects:	Departures due to cohabitation	Departures due to marriage	Departures due to other reasons
a_0 (constant)	-3.329 (0.087)	-3.898 (0.097)	-2.783 (0.065)
a_1 (generation)	1.021 (0.104)	-1.121 (0.051)	0.005 (0.076)
a_2 (generation^2)	-0.98 7 (0.085)	-0.601 (0.089)	-0.061 (0.056)
a_3 (generation^3)	-0.446 (0.150)	-	0.273 (0.104)
a_4 (number of siblings)	0.080 (0.017)	0.087 (0.016)	-
a_5 (foreign father or mother)	-0.489 (0.073)	-	-0.302 (0.058)
a_6 (mother's death)	0.247 (0.115)	-	0.179 (0.096)
a_7 (father farmer)	-0.408 (0.089)	-	-
a_8 (% farmer fathers in département)	-	0.728 (0.342)	1.411 (0.305)
a_9 (previously worked)	0.618 (0.051)	0.810 (0.051)	0.542 (0.042)
a_{10} (previously unemployed)	-	-0.336 (0.115)	-0.483 (0.095)
Random variables, département level :			
σ^2_{u0} (constant)	0.015 (0.010)	0.080 (0.026)	0.020 (0.007)
σ_{u01} (covariance)	-0.019 (0.015)	-	-0.012 (0.007)
σ^2_{u1} (generation)	0.103 (0.043)	-	0.014 (0.013)
σ^2_{u2} (generation^2)	-	0.137 (0.072)	-
σ_{u02} (covariance constant x generation)	-	-0.104 (0.040)	-
Random variable, period level:			
σ^2_{e0}	1	1	1

Again, we see that while all the characteristics significantly influenced all departures, the effect of some characteristics will be confined to certain types of departures. The number of siblings has an identical effect on departures due to cohabitation and marriage, but no effect whatever on departures for other reasons. By contrast, having a foreign father or mother strongly influences departures due to cohabitation or other reasons, but has no effect at all on departures due to marriage, which have a different significance for this category. The same finding applies to the mother's death.

Conclusion

At the end of our presentation of this detailed example of multilevel event-history analysis, we can see the richness of this approach more clearly.

First, it enables us to work on multiple temporalities simultaneously: here, the life course of a generation and a historical time marked by the dates of birth of the generations examined. We could also have introduced the dates of specific events that have marked successive generations at different moments of their lives.

Second, it allows us to operate on several aggregation levels simultaneously: here, we have the individual level, with various characteristics capable of explaining differences in behavior, and the département level, with aggregate characteristics capable of explaining differences observed between départements. Naturally, we could have looked at other aggregation levels. In our example, if we had surveyed different children in the same family, the family level would unquestionably have exerted a significant influence on the children's departure from their family home, as shown by a study conducted in England (Murphy and Wang, 1998).

Just as we distinguished here between departures due to marriage, cohabitation, and other reasons, we could have analyzed simultaneously the occurrence of other phenomena that may be linked to departure from the parental home, such as school-leaving and taking a permanent job. In particular, these models can incorporate random variables to track a heterogeneity that, while unobserved, can nevertheless be identified by observing the occurrence of repetitive events for a given individual, such as births of children and union terminations (Lillard and Waite, 1993). These models also allow the inclusion of clocks and hence multiple temporalities, which make this analysis highly flexible (Lillard, 1993).

General conclusion

The introduction to this book set out its two main objectives: (1) to identify the role played by the opposition between holism and methodological individualism in the history of demographic thought; (2) to attempt to transcend that opposition by replacing it with an approach that synthesizes the various methods proposed earlier while making a fresh contribution to demographic analysis.

The distinction between *holism* and *methodological individualism* has long been discussed in many social sciences (Alexander *et al*. 1987; Berthelot, 2001; Valade, 2001), but its role in demography has not been effectively identified until recently (Courgeau, 2000a, 2003c). It was therefore important to discuss it in detail here, for it sheds light on many problems encountered in the history of demographic thought that echo broader issues in the social sciences.

The opposition between holism and methodological individualism was long regarded as insurmountable, so great was the divergence between their underlying hypotheses and the concepts deriving from them. However, a deeper examination of the two extreme aggregation levels reveals the existence of a multiplicity of other levels, which attenuate the dualist "society versus the individual" approach. Thus it no longer makes sense to choose between holism and individualism, and we need to seek a synthetic approach that we may describe as a *multilevel synthesis*.

In conclusion to this work it is useful to recapitulate the key stages along the path that has led us to this approach and the main hypotheses that it involves. We shall then outline the new avenues of research opened up by the multilevel synthesis.

A historical path from holism to individualism...

Having emerged from the conceptual work of John Graunt in the seventeenth century, demography soon sought to link the main phenomena that it studied—such as death, fertility, and spatial mobility—to the characteristics of the populations that experienced them.

For centuries, until the end of World War II, the dominant approach was what we could call cross-sectional holism, which examined populations at a given moment and tried to explain the aggregate demographic phenomena observed by the general conditions then prevailing. The behavior to be explained was not that of single individuals, but that of a group of individuals (Catholics and Protestants with regard to suicide, for example)—in the belief that social facts originate in the rules governing the society studied, not among the individuals experiencing those facts.

Accordingly, demographers did not need individual data to explain social facts. Methods such as regressions on aggregate data enabled them to identify the sought-for links. Likewise, the introduction of overall indices allowed them to summarize period information with simple measures, which they could also attempt to explain by other characteristics of the population studied.

It is the use of overall period indices that sparked initial doubts about this demographic approach (Henry, 1959), leading to cohort or generational analysis. In the same period, however (Robinson, 1950), sociologists and demographers identified the risk of ecological fallacy, which arises when seeking to explain individual behaviors by aggregate characteristics. The solution—through an event-history approach—did not arrive for some time.

The first point to make here, therefore, is that the period approach, by freezing time at a particular moment without taking account of the human life course, gives an imperfect view of demographic phenomena. This finding and the implementation of resources to overcome it led to the development of a longitudinal holism at the end of World War II. Demographers saw the need to focus their analysis on a generation or cohort in order to correctly capture lifelong phenomena and analyze their history over time.

This makes it possible to assign proper meanings to a phenomenon's intensity, which measures its completeness during the life course of a generation, and its timing, which measures its progress in that lapse of time.

To be applicable, however, this approach requires heroic assumptions about the populations observed. First, for the analysis to be accurate, the population must be homogeneous. Second, only one phenomenon can be studied at a time, under the hypothesis that the other phenomena—called disturbing phenomena—are independent of the phenomenon studied.

In other words, this approach remains holistic, as the homogeneity hypothesis does not allow a separation between the distinct elements of an individual life that could influence the phenomenon studied, while the independence hypothesis eliminates the possible effect of the other demographic phenomena on the progress of the phenomenon studied.

Of course we can always study a phenomenon such as nuptiality in a given sub-population such as farmers. But, if we do this, we shall not always be assured of the homogeneity of the cohort studied, which could well be composed of other, more homogeneous cohorts, and so on. Likewise, new disturbing phenomena, such as exit from agriculture, will join the others and may be even less independent of the phenomenon studied than the other conventional demographic phenomena.

To escape these difficulties, we need to bring individuals into play, with the spectrum of their personal characteristics and, at the same time, their entire life histories, so as to identify the interactions between the different events of their lives. The result is an event-history individualism, which took hold in demography during the 1980s.

The goal of this approach is to identify, in a set of observed life courses, an underlying process that simultaneously involves a set of events experienced by persons during their lives and a set of situations in which people may find themselves. The aim is not to explain an individual life—which lies outside the scope of a social science[13]—but to unravel the

[13] On this topic, see the highly enlightening discussion by Granger (1994) in the chapter "Sur le traitement comme objets des faits humains" ("On the treatment of human facts as objects").

connections between events and situations occurring in a lifetime, which hardly encompass all possible situations.

Thanks to the event-history approach, we can waive the homogeneous-population hypothesis by introducing selected individual characteristics capable of partly explaining the phenomenon(a) analyzed. The approach thus admits the heterogeneity of a population. It enables demographers to set aside the hypothesis of independence between demographic, economic, sociological, and other phenomena while allowing the analysis of interactions between the phenomena.

In other words, the event-history approach shifts the center of interest away from the population (viewed cross-sectionally as an organized whole) or the generation (seen longitudinally as a homogeneous set) to the individual, endowed with specific characteristics that change throughout his or her life. In so doing, however, this methodological individualism exposes itself to the risk of what we call the atomistic fallacy. By ignoring the potential role of society, groups of individuals, and other collective factors on an individual life, the atomistic fallacy contrasts with the ecological fallacy, committed when we restrict the analysis to the effect of society seen as a whole.

...superseded by a multilevel viewpoint

> To begin with, we must consider the fact that between the two extremes—society and the individual—there are many other aggregation levels that must play a role in the evolution of society and the individuals who compose it. In the sphere of interpersonal relationships, the couple, siblings, the family, the household, the contact circle, the neighborhood, family networks, friendly networks, etc. all constitute distinct levels that demand examination in order to understand the role and effect of these networks on human behavior. In the sphere of workplace relationships, the workshop, the firm, the institution, etc. all constitute distinct levels that should be defined more precisely in order to understand their role in the economic evolution of the individual and society. The same goes for political, religious, educational, community-group, and other levels.

This plurality of levels is therefore what stands in the way of an a priori choice between holism and methodological individualism. It is important to study how these levels will connect to and influence one another.

The multilevel approach presented here is an attempt to transcend the divide and supply a method for analyzing these multiple levels simultaneously. We should realize, however, that this method is not necessarily the only alternative or perhaps the best for analyzing such links. Nevertheless, it provides a usable, efficient approach to grasp them.

The first problem encountered when we want to use this method is to identify as best as possible the relevant levels to introduce into a given study. This involves not only a choice among the levels already singled out, but also a search for other levels that are harder to distinguish in a survey but may be important to incorporate in the analysis. Sometimes we shall need to use proxy but not fully relevant segmentations to take account of levels that play a key role in the study. For instance, it seems difficult to use a retrospective survey to capture an individual's life-long networks of relationships. But statistics on the main characteristics of the inhabitants of neighborhoods where the person has lived and worked, for example, already enable us to take fuller account of the types of networks with which (s)he may have been affiliated. Naturally, this is an ersatz, which a prospective survey of individuals over their

lifetimes will be able to replace by their actual affiliation networks. But such a survey would be very unwieldy and difficult to manage.

As well as individual characteristics, the different aggregation levels used in this analysis comprise random variables that must now be clearly defined. The first types of random variables that we can include are linked to the constant term of the individual analysis and characterize the different areas of each aggregation level by the variance that they introduce into the results of the analysis. We therefore set aside the moments of order greater than 2 and assume that the random variables are normally distributed around the value of the constant term.

We can also specify the effect of various characteristics on these random variables, which enables us to generalize the study by making the random variables vary as a function of these characteristics. At the aggregate levels, we have seen the significance we can assign to the random variables: they supply the points (for discrete variables) or curves (for continuous variables) that plot the values of the characteristics in each aggregate area or region examined. They provide a good visual display of the characteristics' effect on the values of the dependent variable, and they highlight the groups or regions where this effect is extreme. We can thus analyze these groups in greater detail to see what distinguishes them from the others, and this may enable us to better explain the outliers.

We can explore another angle by introducing the effects of various characteristics on the individual random variable. This enables us to estimate models that are very different from standard models, in which the random variables are assumed to be normally distributed around a zero mean. Such models are, however, more complex to interpret, as we have shown, because the variances of the random variables estimated for the different characteristics lose the clear significance they had at higher aggregation levels. They can no longer be interpreted as variances of coefficients, but should be viewed more simply as coefficients of quadratic terms describing a more complex function of variance at the individual level.

As seen earlier, these characteristics may have different origins. In particular, they may be taken from the same survey or other surveys performed on the same aggregation levels (Schoumaker, 2001). It may thus be worth incorporating the aggregate characteristic at a given level, corresponding to the individual characteristic measured by the survey on which we are working. We must then detect the aggregate-level variances that we can reduce thanks to this fixed characteristic. This will give us an initial explanation of the differences observed between groups or areas.

While the analysis of random variables allowed us to identify these differences between areas, it hardly enabled us to interpret them more fully. By introducing well-chosen contextual terms, we can reach a clear interpretation of these random variables. For instance, in the example on Norwegian migrations (chapter IV), introducing the percentage of farmers enables us to link the higher probability of migrating by other occupations to the proportion of farmers in a region, providing a clearer explanation of these behaviors.

We then applied these multilevel-analysis methods to the different approaches used in demography, in order to arrive at the analysis that is most relevant and best suited to demographic data: multilevel event-history analysis. It is, however, very interesting to see the application of multilevel analysis to studies on less rich data, of the kind often supplied by ordinary demographic statistics.

For instance, methods for analyzing continuous data are worth considering as some demographic characteristics may present themselves or be summarized in this form. Three

examples are: (1) the DRAT fertility index (Schoumaker and Tabutin, 1999; Schoumaker, 2001), which measures the ratio for each married woman of her actual number of children to the theoretical number of children (given her age and union duration) that she would have had in a natural legitimate fertility profile; (2) the number of dwelling changes for an individual during a period of his or her life; (3) the income of an employee at a given age as a function of his or her initial income and selected characteristics. Using multilevel regression methods poses far more complex problems than a simple regression: risk of attribution to an aggregation level of an individual effect not included in the analysis, attribution of a random effect to an incorrect aggregation level, or the risk of overlooking an aggregation level that should have been examined (Tranmer and Steel, 2001). We have illustrated these problems with specific examples to show how we can try to avoid them.

We have also presented the analysis of discrete data, for more precise event-history data are often lacking in demography. As noted in chapter III, we can introduce individual behaviors even if the analysis does not include duration. This already enables us to clearly separate the effects of many characteristics on the behaviors. The generalization of multilevel methods therefore has its proper place in this work. By applying them to various practical examples, we can identify the pitfalls to be avoided and the different forms that such an analysis can take: modeling of binary data, polytomous data (nominal and ordinal), and event counts.

As noted above, the best approach to demographic phenomena—in all their spatial and temporal scope—is multilevel event-history analysis. However, it poses far more complex problems than the previous forms of analysis, for individuals will not remain in the same area or a given group all their lives. The resulting changes of behavior lead to hypotheses that can be very different and generate divergent behaviors.

For instance, to analyze departure from the parental home, we have used the question on the département of residence immediately prior to the move. We were obliged to use this location as the "Young People and Careers" survey did not give the individual's detailed life history or his or her location(s) throughout childhood. It might have been preferable to select the place where the respondent had spent most of his or her childhood—in the event that it was a different locality than the last place of residence in the parental home—as being a more important factor in studying departure.

These examples show the emergence of the notion that individuals adjust their behaviors to those of the new group that they have just joined. Is the adjustment immediate or fairly rapid or does it require much more time? In the last two cases, how long does it take for the adjustment to occur, and at what pace? Is the adjustment period the same for everyone, or strongly differentiated by individual? All these questions require far richer information than that usually collected by demographic surveys; indeed, they would even call for psychological research. Some data already give us some clues about the adjustment. For instance, the Family survey—carried out as a complement to the 1999 French population census—asked female immigrants to list the date of their settlement in France. The data enable us to identify an effect of their length of stay and their age at migration on the successive births of their children (Toulemon and Mazuy, 2003). We can use this as a measure of their adjustment to host-country behavior. However, such data are still too scarce, and we must stress their importance in conducting a multilevel event-history analysis.

Setting this problem aside, such an analysis proves to be extremely fruitful. It allows us to include several spaces (social, occupational, educational, etc.), whose effect on individual behaviors can be shown. By introducing the characteristics of these spaces, which may vary

over time, we can describe in detail the role of individuals and the social spaces in which they live. These characteristics may, in fact, be obtained from different sources and will enable us to refine the results of the study.

At the same time, we can introduce multiple temporalities (Lillard, 1993), for instance, a historical time with the events occurring in it, which can influence the behaviors studied, and personal temporalities linked to the life histories of the individuals surveyed.

Lastly, we can analyze more complex structures with these methods by introducing, for example, the participation of the same individual in multiple groups: for instance, we can use a single multilevel model to study a given pupil's performance in elementary school and secondary school (Goldstein, 2003).

Probabilities: objectivist, subjectivist or logicist point of view

We must now briefly touch upon some more general issues that closely concern multilevel analysis but cannot be examined in depth in this work, as they would call for broader, more substantive discussion.

First, we should note that multilevel analysis sheds new light on the use of probabilities in social science. The advent of political arithmetic in the seventeenth century—which later gave rise to demography and epidemiology, with the work of John Graunt (1662)—closely follows the introduction of probability theory by Pascal and Fermat (1651). Indeed, political arithmetic was rooted in the principles of probability theory from its very inception. However, while probability theory has been axiomatized for over 70 years now (Kolmogorov, 1933), it is important to realize that many controversies persist over the nature of probability and the scope for inference that they allow. Our aim is not to elaborate on these issues in detail here (Matalon, 1967; Courgeau, 2003d), but to outline the potential role of multilevel analysis in the area.

Throughout our book, we have taken an essentially objectivist attitude—also known as frequentist—to probability. This point of view rests on the law of large numbers established by Jacob Bernoulli (1713), which tells us that the value of the relationship between the number of events observed and the number of individuals at risk tends toward the probability of the event when the population at risk tends toward infinity. Accordingly, when we study the scores obtained by pupils in a given number of schools, we regard this set as a random sample of a supposedly infinite population of schools, which is our objet of study (Goldstein, 2003). It is therefore on this population that we shall conduct our analysis, in terms of mean value and variance, for example.

From this standpoint, statistical inference may allow us to use the observation of the sample as a basis to test the validity of a hypothesis that applies to the population from which the sample is extracted, but not the probability of its being confirmed. To speak of the probability that a hypothesis will be confirmed makes no sense for an objectivist, as we cannot define the probability of an intrinsically unique event.

These conditions seem highly restrictive in social science, where the fact of regarding the population observed as a fraction of a larger population is generally not very realistic. For example, it is hard to consider the population observed in its totality by the Norwegian population register as a sample of a broader population. Moreover, in social science, it seems useful to estimate and compare the probabilities of different hypotheses formulated for a given

population, in order to choose the most likely in the light of the observations. We should therefore turn to a subjectivist attitude toward probabilities.

For this purpose, statistical inference can use the Bayesian approach (1763), which was rejected by the objectivists. The problem addressed by this approach is to determine how the probability of an a priori hypothesis can be modified by later empirical observations. We can easily see why the frequentists are led to reject this statistical inference method, for they refuse to speak of the probability of a hypothesis, which is a unique event. By contrast, for the subjectivist, the notion of the probability of a hypothesis forms the core of probability theory—as they conceive it—and the Bayesian theorem enables them to interpret inference as a decision-making process.

This approach leads them to formulate several axiomatic definitions of "the rational," which we shall not describe in detail here. Readers interested in this approach should read the various books available on Bayesian statistics (Savage, 1954; Lee, 1989; Gelman et al. 1995), which devote several chapters to certain types of multilevel models, also called hierarchical models.

A final approach seeks to give a logicist definition of probabilities. This third point of view shares with the subjectivists the concept of probability as reflecting a degree of belief. For the logicist, however, this degree does not represent a personal feeling, but a logical relationship valid for all, which complements the classical notions of logic. It therefore seeks to assign probabilities with the aid of a logical analysis of the incomplete information encountered in statistical-analysis problems. This approach originated in the work of Laplace (1812), was elaborated by Jeffreys (1939), Cox (1946), and Polya (1954), and was formalized more completely by Jaynes (2003).

There is no room here for a detailed discussion of these issues concerning the definition of probabilities and their application to social science, which lie entirely outside the scope of our work—despite their major role in multilevel analysis (Greenland, 2000). That will be the subject of a separate book, in which we shall clarify the concepts described here in a largely objectivist framework that we can usefully revisit on that occasion.

Toward a more complete theory in social science

As we have already noted, the social sciences came into their own when the events of human life such as birth, illness, and death ceased to be regarded as divine prerogatives off bounds to scientific research, and were recognized as social facts amenable to scientific reasoning.

The complexity of the structures and situations in which a human life unfolds, the role played by people's perception of them and by their adjustment (which differ for every individual), the fact that they are not given once and for all, and that they change from one part of the world to another (through the action of other human beings and the influence of external phenomena)—all these factors have prevented the social sciences from examining humans in their fullness. As a result, there is not one single social science—or "human science"—but many approaches, such as anthropology, demography, economics, psychology, sociology, etc., which try to apportion these very diverse aspects between themselves.

In the first part of our book however, we glimpsed the wider scope of application of event-history analysis and multilevel analysis, which are used in many social sciences. Multilevel analysis has become important in disciplines such as education science (Goldstein,

2003a, 2003b), epidemiology (Morgenstern, 1999; Greenland, 2000; Diez Roux, 1998, 2000, 2003), human geography (Jones, 1997), economics (Walliser, 2003), sociology (DiPrete and Forristal, 1994), and human statistics (Rodriguez and Goldman, 1995; Tranmer et al. 2003). It allows these sciences to transcend some of their major differences and creates an opportunity for a possible convergence between them. To conclude, therefore, let us see examine this new possibility and how it can change our vision of the social sciences.

These sciences developed notions that, from the outset, seemed to be fundamental concepts on which to build all the social sciences. For instance, the object of demography is the quantitative study of human populations through the basic phenomena that drive their change, such as birth, spatial mobility, and death. Likewise, the object of economics is the production, distribution, and consumption of physical goods, through phenomena such as the market. We can easily show that the same pattern applies to all the other social sciences, which seek their object in different aspects of social or individual life.

Now, as Granger (1988) noted:

At this point, we may ask whether multilevel analysis might not be capable of objectivating one of the categories of human experience—by revealing the major role played by aggregation levels on human behaviors and especially the fact that we can no longer consider them separately.

The social sciences initially predicated their identity on the possibility of isolating a series of social objects and individual phenomena, so as to examine them independently. But as the sciences developed, they realized that it was impossible to ignore certain aspects of human life on the grounds that they were totally unrelated to those addressed by another discipline. The analysis in part I of this work shows how demography can no longer view phenomena such as a birth, a migration or a death as being mutually independent and unrelated to the economic, political, educational, and other phenomena that will strongly influence their course. The solution that consists in elaborating an economic, social or political demography offers only a partial solution to the problem, for the challenge is to examine the simultaneous links between these social sciences taken collectively, not two by two. What now needs to be taken into account is the entire society, with its different aggregation levels, and the totality of events, whatever the field where they play a role formerly regarded as predominant.

A far more fundamental point is that, once we have become aware of these links, we shall realize that they connect objects such as fertility, migration, and mortality to the basic phenomena studied previously, such as births, spatial mobility, and deaths. We must now study how the arrangement of units at a lower level (e.g., pupils) explains the properties of the aggregate level (e.g., class or school). The aggregate-level objects are not the sum or the mean of the lower-level objects, for the aggregate level's organization into a whole endows it with new properties. At the same time, we need to study the effect in the opposite direction of the aggregate level on the lower-level units that compose it and that may even, in some cases, entail its elimination. Clearly, this gives a far greater scope to multilevel analysis by leading it to generalize its methods.

Moreover, as already noted, a lower-level unit can, with the help of new units, give birth to different types of higher-level units; between these, there will be interactions requiring analysis. Alternatively, a higher-level unit will, with other units of the same level, form a new structure creating a hierarchy between the levels. This structure of society will not be set for all time; on the contrary, it will evolve continuously.

Lastly, we need to incorporate the physical factors of the environment into this study, without which we cannot study humankind. Of course, they are largely determined by the human groups that shape them, but they can also lie outside their range of action—at least in part, for one can always identify a role of human groups in natural catastrophes.

Our outline of a new approach in social science shows that the multilevel vision lies at the core of all these advances. By allowing social scientists to transcend the opposition between the macro and micro approaches, we have demonstrated that it no longer makes sense to choose between holism and methodological individualism: the task now is to learn how to interconnect the different levels. We have also shown that the notion of research program borrowed from Lakatos (1970) is more effective than Kuhn's notion of paradigm (1962) for situating alternative demographic approaches in time. For instance, if it seems that a theory will be rejected by certain observations, the sensible course is often to protect the program's hard core by reworking other auxiliary hypotheses accordingly: a multilevel approach allows a synthesis between the aggregate and individual approaches, rather than an entirely new theory. By opening up new structures to analysis, it enlarges the field of social sciences while enabling

them to converge, for these more general structures encompass a large number of the sciences. By making it possible to examine simultaneously the multiple significances of a human fact, in a model that incorporates an active temporality, the multilevel approach should bring us closer to the objectivation of human experience and, more generally, toward a new form and a fuller theory in social science—even if the path to this destination remains hard to discern precisely.

Appendix 1

Glossary of epistemological terms[14]

Epistemology: branch of philosophy of science that engages in a critical study of the scientific method and the forms of logic and modes of inference used in science, as well as the principles, basic concepts, theories, and results of individual sciences. Epistemology can thus determine the logical origin, value, and objective significance of these objects of study.

Methodological holism: philosophical doctrine that holds that social phenomena should be studied at their specific level, i.e., the macroscopic level. According to the doctrine, this is a sui generis level of analysis: the true historical "individuals" are not human beings and their idiosyncrasies, but social facts themselves, as well as the supra-individual social actors that produce them, such as institutions, organizations, governments, interest groups, and nations.

Methodological individualism: philosophical doctrine that holds that all concepts and laws specific to the social sciences are reducible to the concepts and laws of individual psychology.

Paradigm: defines the norm of what constitutes legitimate activity within the scientific field that it governs (Kuhn, 1962). The term first denoted the set of beliefs, recognized values, and techniques shared by the members of a particular scientific community. The word can also designate the concrete solutions to enigmas to which members of a given discipline refer in a period of "normal" science. We use the term in the latter sense here, supplementing it with the answer to the question "How do we move from the observed phenomena to the object of the science?"

Research program: a research program rests on a set of basic intuitions and, very often, it does not seem rational to reject these intuitions solely on the basis of incompatible empirical proofs (Lakatos, 1970). In other words, even if a theory seems to be refuted, it often makes sense to protect the program's hard core by reworking the auxiliary hypotheses. Kuhn's "normal" science quite simply describes a research program that enjoys monopoly status. Our opinion is that the social sciences develop in succession by complementing one another rather than by changing paradigms completely.

[14] For further details on these terms, see Nadeau (1999).

Appendix 2

Main software programs suitable for multilevel demographic analysis

The software programs suitable for multilevel analysis fall into two broad categories. The first consists of programs specialized in this type of analysis—often developed by researchers—that allow a wide variety of multilevel approaches. The second comprise more general statistical program banks, featuring some multilevel procedures, often in summary form. The following list is not, of course, exhaustive, but it does offer a choice of possibilities for researchers who want to use multilevel methods. The versions described are those available when this book was published.

I. - More specialized software

aML (Multiprocess Multilevel Modeling): software developed by L.A. Lillard and C.W.A. Panis. Version 2 essentially allows users to estimate event-history models for any number of levels, incorporating a number of simultaneous processes. For more details, see the following web page:
http://www.applied-ml.com.

HLM (Hierarchical Modeling): software developed by S.W. Raudenbush, A.S. Bryk, and R.T. Congdon. Version 5 estimates continuous-response and discrete-response models for 2- or 3-level models only. Cannot handle event-history models. For more details, see the following web page:
http://www.ssicentral.com/hlm/index.html.

MlwiN (Centre for Multilevel Modelling): software developed by J. Rasbash, W. Browne, H. Goldstein, et al. Version 2.1 estimates continuous-response and discrete-response models, event-history models, factor analyses for any given number of levels, and for all possible types of nestings. For more details, see the following web page:
http://www.ioe.ac.uk/mlwin/.

Mplus (Muthén & Muthén): software developed by B. and L. Muthén. Version 2.1 estimates multilevel models for period and generation/cohort analysis for 3 aggregation levels at most. It is actually more specialized in the analysis of structural linear models with latent variables, such as LISREL. For more details, see the following web page:
http://www.statmodel.com/features.shtml.

II. - More general software

BMDP: Version 7.0 of this program bank only allows the analysis of variance components for multilevel data. For more details, see the following web page:
http://www.statsol.ie/html/bmdp/bmdp_features.html#a10.

GenStat: Version 7.1 of this program bank estimates continuous-response and discrete-response models, for different possible types of nestings. For more details, see the following web page:
http://www.vsni.co.uk/products/genstat/.

SAS: Version 9.1 of this program bank estimates continuous-response and discrete-response models for 2 aggregation levels at most and for different possible types of nestings (NLMIXED procedure). For more details, see the following web page:
http://support.sas.com/software/index.htm.

S-PLUS: Version 6 of this program bank estimates continuous-response and discrete-response models, for different possible types of nestings but only 2 aggregation levels. For more details, see the following web page:
http://www.insightful.com/.

SPSS: Version 12 of this program bank estimates continuous-response and discrete-response models for 2 aggregation levels at most. For more details, see the following web page:
http://www.spss.com/spss/.

Stata: Version 8 of this program bank estimates continuous-response and discrete-response models for 2 aggregation levels at most. For more details, see the following web page:
http://www.stata.com/.

Bibliography

Aitkin, M., Bennett, S.N., and Hesketh, J. (1981), Teaching styles and pupil progress: a reanalysis, British Journal of Educational Psychology, 51, pp. 170-186.

Alexander, M., Giesen, B., Münch, R., and Smelser, N.J. (eds) (1987), The micro-macro link, University of California Press, Berkeley, Los Angeles, London.

Alker, H.R. (1969), A typology of ecological fallacies, in M. Dogan and S. Rokkan (eds), Quantitative ecological analysis, MIT Press, Cambridge (MA), pp. 69-90.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993), Statistical models based on counting processes, Springer-Verlag, New York.

Aristotle [1885], Politics, transl. by B. Jowett, http://classics.mit.edu/Aristotle/politics.mb.txt.

Aristotle [1954], Rhetoric, transl. by W. Rhys Roberts, http://classics.mit.edu/Aristotle/rhetoric.mb.txt.

Aubin, J.P. (1991), Viability theory, Birkäuser, Boston, Basel, Berlin.

Auriat, N. (1996), Les défaillances de la mémoire humaine. Aspects cognitifs des enquêtes rétrospectives, Travaux et Documents, 136, INED, Paris.

Baccaïni, B. and Courgeau, D. (1996), Approche individuelle et approche agrégée: utilisation du registre de population norvégien pour l'étude des migrations, in J.P. Bocquet-Appel, D. Courgeau, and D. Pumain (eds), Analyse spatiale de données biodémographiques / Spatial analysis of biodemographic data, Congresses and Colloquia, 16, INED / John Libbey Eurotext, Montrouge, pp. 79-104.

Bagdonavicius, V. and Nikulin, M. (2002), Accelerated life models, Chapman and Hall, London.

Bayes, T.R. (1763), An essay towards solving a problem in the doctrines of chance, Philosophical Transactions of the Royal Society of London, 53, pp. 370-418.

Becker, G.S. (1960), An economic analysis of fertility, in Demographic and economic change in developed countries: A Conference of the Universities–National Bureau Committee for Economic Research, Princeton University Press, Princeton.

Bennett, S.N. (1976), Teaching styles and pupil progress, Open Book, London.

Bernoulli, J.I. (1713), Ars conjectandi, Impensis Thurnisiorum Fratrum, Basel.

Berthelot, J.M. (2001), Les sciences du social, in J.M. Berthelot (ed.), Epistémologie des sciences sociales, Presses Universitaires de France (PUF), Paris, pp. 203-265.

Berthelot, J.M. (ed.) (2001), Epistémologie des sciences sociales, Presses Universitaires de France (PUF), Paris.

Birnbaum, P. and Leca, D. (eds) (1986), Sur l'individualisme. Thèmes et méthodes, Presses de la Fondation Nationale des Sciences Politiques, Paris.

Blayo, C. (1995), La condition d'homogénéité en analyse démographique et en analyse statistique des biographies, Population, 50, 6, pp. 1501-1518.

Bonneuil, N. (1994), Capital accumulation, inertia of consumption and norms of reproduction, Journal of Population Economics, 7, pp. 49-62.

Bonneuil, N. (1997), Jeux, équilibres et régulation des populations sous contrainte de viabilité. Une lecture de l'œuvre de l'anthropologue Fredrik Barth, in D. Courgeau (ed.), Nouvelles approches méthodologiques en sciences sociales, Population, 52, 4, pp. 947-976. English ed.: (1998), Games, equilibria and population regulation under viability constraints: An interpretation of the work of the anthropologist Fredrik Barth, in D. Courgeau (ed.), New methodological approaches in the social sciences, Population: an English Selection, 10, 1, pp. 151-179.

Boudon, R. (1977), Effets pervers et ordre social, Presses Universitaires de France (PUF), Paris.

Boudon, R. (1988), Individualisme ou holisme: un débat méthodologique fondamental, in H. Mendras and M. Verret (eds), Les champs de la sociologie française, Armand Colin, Paris, pp. 31-45.

Box, G.E.P. and Cox, D.R. (1964), An analysis of transformations, Journal of the Royal Statistical Society B, 26, pp. 211-252.

Brémaud, P. and Jacod, J. (1977), Processus ponctuels et martingales: résultats récents sur la modélisation et le filtrage, Advanced Applied Probabilities, 9, pp. 362-416.

Breslau, N.E. and Clayton, D.G. (1993), Approximate inferences in generalized linear models, Journal of the American Statistical Association, 88, 421, pp. 9-25.

Bretagnole, J. and Huber-Carol, C. (1988), Effects of omitting covariates in Cox's model for survival data, Scandinavian Journal of Statistics, 15, pp. 125-138.

Browne, N.J., Goldstein, H., and Rasbash, J. (2001), Multiple membership multiple classification (MMMC) models, Statistical modelling, 1, pp. 103-124.

Bry, X. (1996), Analyses factorielles multiples, Economica, Paris.

Bryk, A.S. and Raudenbush, S.W. (1992), Hierarchical linear models: applications and data analysis, Sage, Newbury Park (CA).

Burch, T.K. (1999), Something ventured, something gained: Progress towards a unified theory of fertility decline, in D. Tabutin, C. Gourbin, G. Masuy-Strobant, and B. Schoumaker (eds), Théories, paradigmes et courants explicatifs en démographie, Chaire Quetelet 1997, Academia-Bruylant / L'Harmattan, Louvain-la-Neuve, pp. 253-278.

Burch, T.K. (2002), Computer modelling of theory: explanations for the 21th century, in R. Franck (ed.), The explanatory power of models, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 245-266.

Caldwell, J.C. (1982), Theory of fertility decline, Academic Press, London.

Caselli, G., Vallin, J., and Wunsch, G. (eds) (2001), Démographie: analyse et synthèse, I, La dynamique des populations, INED, Paris. English ed.: (2005), Demography: analysis and synthesis. A treatise in population, 4 vols, Amsterdam, Oxford, Academic Press (Elsevier).

Chabanne, A. and Lollivier, S. (1988), Les salariés de 1967, quinze ans après: la trace du chemin parcouru, Économie et Statistique, 210, pp. 21-32.

Clayton, D. and Cuzick, J. (1985), Multivariate generalizations of the proportional hazard model, Journal of the Royal Statistical Society A, 148, pp. 82-117.

Cleland, J. (1985), Marital fertility decline in developing countries: Theories and the evidence, in J. Cleland and J. Hobcraft (eds), Reproductive change in developing countries: Insights from the World Fertility Survey, Oxford University Press, Oxford, pp. 223-252.

Coale, A.J. (1973), The demographic transition reconsidered, in International Population Conference, Liège, vol. 1, IUSSP, Liège, pp. 53-72.

Commenges, D. (1999), Multistate models in epidemiology, Lifetime Data Analysis, 5, pp. 315-327.

Condorcet, M.J.A.N. Caritat, Marquis de (1994), Arithmétique politique. Textes rares ou inédits (1767-1789), critical ed. annotated by B. Bru and P. Crépel, INED / Presses Universitaires de France (PUF) Diffusion, Paris.

Courgeau, D. (1973), Migrants et migrations, Population, 28, 1, pp. 95-129.

Courgeau, D. (1982), Premiers migrants, migrants secondaires et retours (France 1968-1975), Population, 37, 6, pp. 1189-1193.

Courgeau, D. (1987), Constitution de la famille et urbanisation, Population, 42, 1, pp. 57-82. English ed.: (1989), Family Formation and Urbanization, Population: an English Selection, 1, pp. 123-146.

Courgeau, D. (1988), Méthodes de mesure de la mobilité spatiale. Migrations internes, mobilité temporaire, navettes, INED, Paris.

Courgeau, D. (1991a), Perspectives avec migrations, Population, 46, 6, pp. 1513-1530.

Courgeau, D. (1991b), Analyse de données biographiques érronées, Population, 46, 1, pp. 89-104. English ed.: (1992), Impact of response errors on event history analysis, Population: an English Selection, 4, pp. 97-110.

Courgeau, D. (1999a), L'enquête "Triple biographie: familiale, professionnelle et migratoire," in Groupe de Réflexion sur l'Approche Biographique, Biographies d'enquêtes, "Méthodes et Savoirs" series, INED / Presses Universitaires de France (PUF) Diffusion, Paris, pp. 59-74.

Courgeau, D. (1999b), De l'intérêt des analyses multi-niveaux pour l'explication en démographie, in D. Tabutin, C. Gourbin, G. Masuy-Strobant, and B. Schoumaker (eds), Théories, paradigmes et courants explicatifs en démographie, Chaire Quetelet 1997, Academia-Bruylant / L'Harmattan, Louvain-la-Neuve, pp. 93-116.

Courgeau, D. (2000a), Réflexions sur la causalité en sciences sociales, Recherches et prévisions, 60, pp. 49-60.

Courgeau, D. (2000b), Le départ de chez les parents: une analyse démographique sur le long terme, Économie et Statistique, 337-338, pp. 37-60.

Courgeau, D. (2001a), Multi-state transition models in demography, in J. Hoem (ed.), International Encyclopedia of the Social and Behavioral Sciences, 15, Demography, Pergamon, Oxford, pp. 10210-10214.

Courgeau, D. (2001b), Individus et contextes dans l'analyse des comportements selon l'approche multiniveau, in G. Caselli, J. Vallin and G. Wunsch (eds), Démographie: analyse et synthèse, I, La dynamique des populations, INED, Paris, pp. 519-536. English ed.: (2005), Demography: analysis and synthesis. A treatise in population, 4 vols, Amsterdam, Oxford, Academic Press (Elsevier).

Courgeau, D. (2002a), Vers une analyse biographique multiniveau, in Actes des Journées de Méthodologie Statistique, 2, INSEE Méthodes, 101, Paris, pp. 375-394.

Courgeau, D. (2002b), Évolution ou révolution dans la pensée démographique, Mathématiques et Sciences Humaines, 160, pp. 49-78.

Courgeau, D. (2002c), New approaches and methodological innovations in the study of partnership and fertility behaviour, in M. Macura and G. Beets (eds), Dynamics of fertility and partnership in Europe. Insights and lessons from comparative research, United Nations, Geneva, pp. 99-114.

Courgeau, D. (ed.) (2003a), Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London.

Courgeau, D. (2003b), General introduction, in Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp.1-24.

Courgeau, D. (2003c), From the macro-micro opposition to multilevel analysis, in Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 43-92.

Courgeau, D. (2003d), General conclusion, in Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 199-214.

Courgeau, D. and Baccaïni, B. (1997), Analyse multi-niveaux en sciences sociales, in D. Courgeau (ed.), Nouvelles approches méthodologiques en sciences sociales, Population, 52, 4, pp. 831-864. English ed.: (1998), Multilevel analysis in the social sciences, in D. Courgeau (ed.), New methodological approaches in the social sciences, Population: an English Selection, 10, 1, pp. 39-71.

Courgeau, D. and Lelièvre, É. (1985), Nuptialité et agriculture, Population, 41, 2, pp. 303-326.

Courgeau, D. and Lelièvre, É. (1989), Analyse démographique des biographies, INED, Paris.

Courgeau, D. and Lelièvre, É. (1992), Event history analysis in demography, Clarendon Press, Oxford.

Courgeau, D. and Lelièvre, É. (1996), Changement de paradigme en démographie, Population, 2, 51, pp. 645-654. English ed.: (1997), Changing paradigm in demography, Population: an English Selection, 9, pp. 1-10.

Courgeau, D. and Lelièvre, É. (2001), Análisis demográfico de las biografías, El Collegio de Mexico, Mexico City.

Courgeau, D. and Meron, M. (1996), Trajectoires d'activité des couples, in A. Degenne, M. Mansuy, G. Podevin, and P. Werkin (eds), Typologie des marchés du travail: suivi et parcours, 115, CEREQ, pp. 239-255.

Cox, D.R. (1970), The analysis of binary data, Chapman and Hall, London.

Cox, D.R. (1972), Regression models and life tables (with discussion), Journal of the Royal Statistical Society B, 34, pp. 269-276.

Cox, D.R. and Hinkley, D.V. (1974), Theoretical statistics, Chapman and Hall, London.

Cox, R. (1946), Probability, frequency, and reasonable expectation, American Journal of Physics, 14, pp. 1-13.

Cribier, F. and Kych, A. (1999), Un ensemble d'enquêtes auprès de deux cohortes de retraités parisiens, in Groupe de Réflexion sur l'Approche Biographique, Biographies d'enquêtes, "Méthodes et Savoirs" series, INED / Presses Universitaires de France (PUF) Diffusion, Paris, pp. 75-103.

Davis, K. (1945), The world demographic transition, Annals of the American Academy of Political and Social Science, 237, pp. 1-11.

De Finetti, B. (1974), Theory of probability, 2 vols, Wiley & Sons, London, New York.

Degenne, A. and Forsé, M. (1994), Les réseaux sociaux, Armand Colin, Paris.

Degenne, A. and Forsé, M. (1999), Introducing social networks, Sage Publication, London.

Delaporte, P. (1941), Évolution de la mortalité en Europe depuis les origines des statistiques de l'état civil, Paris.

Delaunay, D. (2001), L'inscription dans l'espace des biographies individuelles, IUSSP General Population Conference, Salvador (Brazil).

Dellacherie, C. (1980), Un survol de l'intégrale stochastique, Stochastic Processes Applications, 10, pp. 115-144.

Dellacherie, C. and Meyer, P.A. (1980), Probabilités et potentiels: théorie des martingales, Hermann, Paris.

Deparcieux, A. (1746), Essai sur les probabilités de la durée de la vie humaine, Guérin Frères, Paris.

Deparcieux, A. (2003), Essai sur les probabilités de la durée de la vie humaine (1746) followed by Addition à l'Essai (1760), facsimile reprint with introduction and notes by C. Behar and contributions by G. Gallais-Hamono, C. Rietsch, and J. Berthon, INED, Paris.

Depardieu, D. (1978), Disparités de salaire dans le tertiaire, Économie et Statistique, 98, pp. 21-30.

Depardieu, D. and Payen, J.F. (1986), Disparités de salaire dans l'industrie en France et en Allemagne: des ressemblances frappantes, Économie et Statistique, 188, pp. 23-34.

Desplanques, G. (1984), L'inégalité sociale devant la mort, Économie et Statistique, 162, pp. 29-50.

Diez Roux, A.V. (2003), Potentialities and limitations of multilevel analysis in public health and epidemiology, in D. Courgeau (ed.), Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 93-120.

Diggle, P.J. and Liang, K.Y. (1994), Analysis of longitudinal data, Clarendon Press, Oxford.

DiPrete, T. and Forristal, J. (1994), Multilevel analysis: methods and substance, Annual Review of Sociology, 24, pp. 331-357.

Duchêne, J., Wunsch G., and Vilquin, É. (eds) (1989), L'explication en sciences sociales: la recherche des causes en démographie, Ciaco, Brussels.

Duncan, C. (1997), Applying multivariate multilevel models in geographical research, in G.P. Westert and R.N. Verhoeff (eds), Places and people: multilevel modelling in geographical research, Nederlandse Geografische Studies, 227, Urban Research Center, Utrecht, pp. 100-119.

Duncan, O. D., Cuzzort, R.P., and Duncan, B. (1961), Statistical Geography: Problems in analysing areal data, Free Press, Glencoe.

Dupâquier, J. (1985), Le mémoire de Jean de Witt sur la valeur des rentes viagères, Annales de démographie historique, pp. 355-394.

Dupâquier, J. (1996), L'invention de la table de mortalité, Presses Universitaires de France (PUF), Paris.

Dupâquier, J. and M. (1985), L'histoire de la démographie, Librairie Académique Perrin, Paris.

Durkheim, É. (1930), Le suicide, Presses Universitaires de France (PUF), Paris (1st ed. 1897, Félix Alcan, Paris).

Durkheim, É. (1937), Les règles de la méthode sociologique, Presses Universitaires de France (PUF), Paris (1st ed. 1895, Félix Alcan, Paris).

Easterlin, R.A. (1961), The American baby-boom in historical perspective, American Economic Review, 51, pp. 860-911.

Easterlin, R.A. and Crimmins, E.N. (1985), The fertility revolution: A demand-supply analysis, University of Chicago Press, Chicago.

Euler, L. (1760), Recherches générales sur la mortalité et la multiplication du genre humain, Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Berlin, VII, pp. 144-175.

Fielding, A., Yang M, and Goldstein, H. (2003), Multilevel ordinal models for examination grades, Statistical Modelling, 3, 2, pp. 127-153.

Filloux, J.C. (2001), Épistémologie, éthique et sciences de l'éducation, L'Harmattan, Paris.

Firebaugh, G. (1978), A rule for inferring individual-level relationships from aggregate data, American Sociological Review, 43, pp. 557-572.

Franck, R. (ed.) (1994), Faut-il chercher aux causes une raison? L'explication dans les sciences humaines, Librairie Philosophique Vrin, Paris.

Franck, R. (1995), Mosaïques, machines, organismes et sociétés. Examen metadisciplinaire du réductionnisme, Revue Philosophique de Louvain, 93, pp. 67-81.

Franck, R. (ed.) (2002), The explanatory power of models, Kluwer Academic Publishers, Boston, Dordrecht, London.

Franck, R. (2003), Causal analysis, systems analysis, and multilevel analysis: Philosophy and epistemology, in D. Courgeau (ed.), Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 175-198.

Gelman, A., Karlin, J.B., Stern, H.S., and Rubin, D.B. (1995), Bayesian data analysis, Chapman and Hall, New York.

Giddens, A. (1984), The constitution of society, Polity Press, Cambridge (UK).

Glaude, M. and Jarousse, J.P. (1988), L'horizon des jeunes salariés dans leur entreprise: du lien noué avec l'employeur dépend en partie le salaire et la carrière, Économie et Statistique, 211, pp. 23-41.

Goldstein, H. (1986), Multilevel mixed linear model analysis using iterative generalized least squares, Biometrika, 73, pp. 43-56.

Goldstein, H. (1987), Multilevel covariance component models, Biometrika, 74, pp. 430-431.

Goldstein, H. (1989), Restricted unbiased generalized least-squares estimation, Biometrika, 76, pp. 622-623.

Goldstein, H. (1991), Nonlinear multilevel models, with an application to discrete response data, Biometrika, 78, pp. 45-51.

Goldstein, H. (2003a), Multilevel statistical models, Arnold, London (1st ed. 1987 as Multilevel models in educational and social research).

Goldstein, H. (2003b), Multilevel modelling of educational data, in D. Courgeau (ed.), Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 25-42.

Goldstein, H. and Rasbash, J. (1996), Improved approximations for multilevel models with binary responses, Journal of the Royal Statistical Society A, 159, 3, pp. 505-513.

Goldstein, H., Rabash, J., Browne, W., Woodhouse, G., and Poulain, M. (2000), Models in the study of dynamic household structures, European Journal of Population, 16, 4, pp. 373-387.

Gourieroux, C. (1989), Économétrie des variables qualitatives, Economica, Paris.

Granger, G.G. (1988), Essai d'une philosophie du style, Odile Jacob, Paris.

Granger, G.G. (1994), Formes, opérations, objets, Librairie Philosophique J. Vrin, Paris.

Granger, G.G. (2001), Sciences et réalités, Odile Jacob, Paris.

Graunt, J. (1662), Natural and political observations upon the bills of mortality [...] of the city of London, Tho. Roycroft, London. French ed.: (1977), Observations naturelles et politiques, transl. by É. Vilquin, INED, Paris.

Greenland, S. (1998a), Hierarchical regression, in K. Rothman and S. Greenland (eds), Modern Epidemiology, Lippincott-Raven, Philadelphia, pp. 427-432.

Greenland, S. (1998b), Probability logic and probability induction, Epidemiology, 9, pp. 322-332.

Greenland, S. (2000), Principles of multilevel modelling, International Journal of Epidemiology, 29, pp. 158-167.

Groupe de Réflexion sur l'Approche Biographique (1999), Biographies d'enquêtes, "Méthodes et Savoirs" series, INED / Presses Universitaires de France (PUF) Diffusion, Paris.

Guillard, A. (1855), Éléments de statistique humaine ou démographie comparée, Guillaumin, Paris.

Gumpertz, M.L. and Pantula, S.G. (1992), Nonlinear regression with variance components, Journal of the American Statistical Association, 87, 417, pp. 201-209.

Halley, E. (1693), An estimate of the degrees of the mortality of mankind drawn from curious tables of the births and funerals at the city of Breslaw, Philosophical Transactions, XVII, pp. 596-610.

Hartley, H.O. and Rao, J.N.K. (1967), Maximum likelihood estimation for the mixed analysis of variance model, Biometrika, 54, pp. 93-108.

Harville, D. (1976), Extension of the Gauss-Markov theorem to include the estimation of random effects, Annals of Statistics, 4, 2, pp. 384-395.

Henry, L. (1959), D'un problème fondamental d'analyse démographique, Population, 13, 1, pp. 9-32.

Henry, L. (1966), Analyse et mesure des phénomènes démographiques par cohorte, Population, 20, 3, pp. 465-482.

Henry, L. (1972), Démographie: Analyse et modèles, Larousse, Paris.

Henry, L. (1981), Dictionnaire démographique multilingue, UIESP / Ordina Editions, Liège. English section: (1982), Multilingual Demographic Dictionary, 2nd ed., adapted by É. van de Walle, IUSSP / Ordina Editions, Liège.

Hoem, J. (1983), Multistate mathematical demography should adopt the notions of event-history analysis, Stockholm Research Reports in Demography, 10.

Hoem, J. (1985), Weighting, misclassification and other issues in the analysis of survey samples of life histories, in J. Heckman and B. Singer (eds), Longitudinal analysis of labour market data, Cambridge University Press, Cambridge (UK), pp. 249-293.

Hougaard, P. (1986), Survival models for heterogeneous populations derived from stable distributions, Biometrika, 73, pp. 387-396.

Hougaard, P. (2000), Analysis of multivariate survival data, Springer-Verlag, New York, Berlin, Heidelberg.

Hougaard, P., Harvald, B., and Holm, N.V. (1992), Measuring the similarity between the lifetimes of adult Danish twins born between 1881-1930, Journal of the American Statistical Association, 87, pp. 17-24.

Hubern J. (ed.) (1991), Macro-micro linkages in sociology, Sage, Newbury Park (CA).

Jaynes, E.T. (2003), Probability theory. The logic of science, G.L. Bretthorst (ed.), Cambridge University Press, Cambridge (UK), New York, Melbourne.

Jeffreys, H. (1939), Theory of probability, Clarendon Press, New York.

Jones, K. (1997), Multilevel approaches to modelling contextuality: from nuisance to substance in the analysis of voting behaviour, in G.P. Westert and R.N. Verhoeff (eds), Places and people: multilevel modelling in geographical research, Nederlandse Geografische Studies, 227, Urban Research Center, Utrecht, pp. 19-43.

Kalbfleisch, J.D. and Prentice, R. (1973), Marginal likelihood based on Cox's regression and life model, Biometrika, 60, pp. 267-278.

Keiding, N. (1999), Event history analysis and inference from observational epidemiology, Statistics in Medicine, 18, pp. 2353-2363.

Klein, J.P. (1992), Semiparametric estimation of random effects using the Cox model based on the EM algorithm, Biometrics, 48, pp.795-806.

Kolmogorov, A.N. (1933), Grundbegriffe der Warhrscheinlichkeitsrenung, in Ergebnisse der Mathematik, 2, Springer, Berlin.

Körösi, J. (1894), An estimate of the degree of legitimate natality, as shown in the table of natality compiled by the author from observations made at Budapest, Proceedings of the Royal Society of London, 55, 331, pp. 16-17.

Kreft, I., and de Leeuw, J. (1998), Introducing multilevel modelling, Sage, London.

Kuhn, T.S. (1962), The structure of scientific revolutions, University of Chicago Press, Chicago.

Lakatos, I. (1970), Falsification and methodology of scientific research programmes, in I. Lakatos and A. Musgrave (eds), Criticism and the growth of knowledge, Cambridge University Press, Cambridge (UK), pp. 91-196.

Landry, A. (1934), Les trois théories principales de la population, in La révolution démographique, Sirey, Paris (1st ed. 1909).

Landry, A. (1945), Traité de démographie, Payot, Paris.

Laplace, P.S., Marquis de (1812), Théorie analytique des probabilités, 2 vols, Coursier Imprimeur, Paris.

Lazarsfeld, P.F. and Menzel, H. (1961), On the relation between individual and collective properties, in A. Etzioni (ed.), Complex organizations, Holt, Reinhart, and Winston, New York, pp. 422-440.

Lee, P.M. (1997), Bayesian statistics. An introduction, 2nd ed., Arnold, London.

Lelièvre, É. (1987), Activité professionnelle et fécondité: les choix et les déterminations des femmes françaises entre 1930 et 1960, Cahiers Québéquois de Démographie, 16, 2, pp. 207-236.

Lelièvre, É., Bonvalet, C., and Bry, X. (1997), Analyse biographique des groupes. Les avancées d'une recherche en cours, in D. Courgeau (ed.), Nouvelles approches méthodologiques en sciences sociales, Population, 52, 4, pp. 803-830. English ed.: (1998), Event history analysis of groups. The findings of an on-going project, in D. Courgeau (ed.), New methodological approaches in the social sciences, Population: an English Selection, 10, 1, pp. 11-37.

Lelièvre, É. and Bringé, A. (1998), Manuel pratique pour l'analyse statistique des biographies / Practical Guide to Event History Analysis, "Méthodes et Savoirs" series, INED / Presses Universitaires de France (PUF), Paris.

Lesourne, J. (1991), The economics of order and disorder, Clarendon Press, Oxford.

Lesthaeghe, R. (1983), A century of demographic and cultural change in Western Europe: An exploration of underlying dimensions, Population and Development Review, 9, 3, pp. 411-435.

Lillard, L.A. (1993), Simultaneous equations for hazards: Marriage duration and fertility timing, Journal of Econometrics, 56, pp. 189-217.

Lillard, L.A. and Waite, L.J. (1993), A joint model of marital childbearing and marital disruption, Demography, 30, 4, pp. 653-681.

Lindsey, J.K. (1999), Models for repeated measurements, Oxford University Press, Oxford.

Lindstrom, M.J. and Bates, D.M. (1990), Nonlinear mixed effects models for repeated measures data, Biometrics, 46, pp. 673-687.

Longford, N.T. (1987), A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects, Biometrika, 74, 4, pp. 817-827.

Loriaux, M. (1989), L'analyse contextuelle: renouveau théorique ou impasse méthodologique, in J. Duchêne, G. Wunsch, and É. Vilquin (eds), L'explication en sciences sociales. La recherche des causes en démographie, Chaire Quetelet 1987, Ciaco, Brussels, pp. 333-368.

Lotka, A. (1939), Théorie analytique des associations biologiques, part 2, Hermann, Paris.

Lyberg, I. (1983), The effect of sampling and nonresponse on estimates of transition intensities: some empirical results from the 1981 Swedish Fertility Survey, Stockholm Research Reports in Demography, 14.

Ma, R., Krewski, D., and Burnett, R.T. (2003), Random effects Cox model: A Poisson modelling approach, Biometrika, 90, 1, pp. 157-169.

Manton, K.G., Singer, B., and Woodbury, M.A. (1992), Some issues in the quantitative characterization of heterogeneous populations, in Demographic applications of event history analysis, Clarendon Press, Oxford, pp. 9-37.

Mason, K.O. (1997), Explaining fertility transitions, Demography, 34, 4, pp. 443-454.

Mason, W.M., Wong, G.W., and Entwistle, B. (1983), Contextual analysis through the multilevel linear model, in S. Leinhart (ed.), Sociological Methodology 1983-1984, Jossey-Bass, San Francisco, pp. 72-103.

Matalon, B. (1967), Épistémologie des probabilités, in J. Piaget (ed.) Logique et connaissance scientifique, Encyclopédie de la Pléiade, Gallimard, Paris, pp. 526-553.

Matsuyama, Y., Sakamoto, J., and Ohashi, Y. (1998), A Bayesian hierarchical survival model for the institutional effects in a multi-centre cancer clinical trial, Statistics in Medicine, 17, pp. 1893-1908.

McCullagh, P. (1980), Regression models for ordinal data, Journal of the Royal Statistical Society B, 42, 2, pp.109-142.

McCullagh, P. and Nelder, J.A. (1983), Generalized linear models, Chapman and Hall, London.

McMichael, A.J. (1999), Prisoners of the proximate: loosening the constraints on epidemiology in an age of change, American Journal of Epidemiology, 149, pp. 887-897.

Menken, J. and Trussel, J. (1981), Proportional hazards life table models: An illustrative analysis of socio-demographic influences on marriage dissolution in the United States, Demography, 18, 2, pp. 181-200.

Miller, J.J. (1977), Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance, Annals of Statistics, 5, 4, pp. 746-762.

Moheau, J.B. (1778), Recherches et considérations sur la population de la France, Paris (1912 reprint ed. by R. Gonnard, Geuthner, Paris).

Moheau, J.B. (1994), Recherches et considérations sur la population de la France, reprint with annotations by É. Vilquin, INED, Paris.

Morgenstern, H. (1998), Multilevel analyses and design, in K. Rothman and S. Greenland (eds), Modern Epidemiology, Lippincott-Raven, Philadephia, pp. 478-480.

Murphy, M. and Wang, D. (1998), Family and sociodemographic influences on patterns of leaving home in postwar Britain, Demography, 35, 3, pp. 293-305.

Nadeau, R. (1999), Vocabulaire technique et analytique de l'épistémologie, Presses Universitaires de France (PUF), Paris.

Notestein, F. (1945), Population - The long view, in T.M. Schulz (ed.), Food in the world, University of Chicago Press, Chicago.

Pascal, B. (1651), Traité du triangle arithmétique, in A.W.F. Edwards (ed.), Pascal's arithmetic triangle, Charles Griffin, London, 1986.

Peto, M. and R. (1972), Asymptotically efficient rank invariant test procedures (with discussion), Journal of the Royal Statistical Society A, 135, pp. 185-206.

Petty, W. (1690), Political arithmetick, London (1963 reprint in C.H. Hull (ed.), Reprints of economic classics, vol. 1, A.M. Kelley, New York, pp. 121-231).

Piaget, J. (ed.) (1967), Logique et connaissance scientifique, Encyclopédie de la Pléiade, Gallimard, Paris.

Polya, G. (1954), Mathematics and plausible reasoning, 2 vols, Princeton University Press, Princeton.

Popper, K.R. (1959), The logic of scientific discovery, Hutchinson, London.

Poulain, M. (1996), Le registre de population centralisé: un excellent outil de mesure multi-niveau, in J.P. Bocquet-Appel, D. Courgeau, and D. Pumain (eds), Analyse spatiale de données biodémographiques / Spatial analysis of biodemographic data, Congresses and Colloquia, 16, INED / John Libbey Eurotext, Montrouge, pp. 63-77.

Poulain, M., Riandey, B., and Firdion, J.M. (1991), Enquête biographique et registre belge de population: une confrontation des données, Population, 46, 1, pp. 65-88. English ed.: (1992), Data from a life history survey and the Belgian Population Register: a comparison, Population: an English Selection, 4, pp. 77-96.

Pourcher, G. (1966), Un essai d'analyse par cohorte de la mobilité géographique et professionnelle, Population, 21, 2, pp. 357-378.

Pressat, R. (1966), Principes d'analyse, INED, Paris.

Puig, J.P. (1981), La migration régionale de la population active, Annales de l'INSEE, 44, pp. 41-79.

Pumain, D. and Saint-Julien, T. (1990), France, in R. Brunet (ed.), France, Europe du Sud, Hachette / Reclus, Paris.

Quetelet, A. (1869), Physique sociale ou essai sur le développement des facultés de l'homme, Muquardt, Brussels; Baillière et Fils, Paris; Issakoff, Saint Petersburg (1997 reprint by Académie Royale de Belgique).

Raftery, A.E. (1995), Bayesian model selection in social research, Sociological Methodology, 25, pp. 111-163.

Raudenbush, S.W. and Bryk, A.S. (1986), A hierarchical model for studying school effects, Sociology of Education, 59, pp. 1-17.

Rice, N. and Leyland, A. (1996), Multilevel models: Applications to health data, Journal of Health Services Research & Policy, 1, pp. 154-164.

Robinson, W.S. (1950), Ecological correlations and the behavior of individuals, American Sociological Review, 15, pp. 351-357.

Rodriguez, G. and Goldman, N. (1995), An assessment of estimation procedures for multilevel models with binary responses, Journal of the Royal Statistical Society A, 158, 1, pp. 73-89.

Rogers, A. (1973), The multiregional life table, Journal of Mathematical Sociology, 3, pp. 127-137.

Roussel, L. (1971), La nuptialité en France. Précocité et intensité suivant les régions et les catégories socio-professionnelles, Population, 26, 6, pp. 1029-1056.

Ryder, N.B. (1965), The cohort as a concept in the study of social change, American Sociological Review, 30, 6, pp. 843-861.

Sadler, M.T. (1830), The law of population: a treatise, in six books, in disproof of the superfecundity of human beings, and developing the real principle of their increase, John Murray, London.

Sargent, D.J. (1998), A general framework for random effects survival analysis in the Cox proportional hazards setting, Biometrics, 54, pp. 1486-1497.

Sastry, N. (1997), Family-level clustering of childhood mortality in Northeast Brazil, Population Studies, 51, pp. 245-261.

Savage, L.J. (1954), The foundations of statistics, Wiley, New York.

Schall, R. (1991), Estimation in generalized linear models with random effects, Biometrika, 78, 4, pp. 719-727.

Schoen, R. and Nelson, V.E. (1974), Marriage, divorce, and mortality: a life table analysis, Demography, 11, pp. 267-290.

Schoumaker, B. (1999), Analyse multi-niveaux et explication de la fécondité dans les pays du Sud, in D. Tabutin, C. Gourbin, G. Masuy-Strobant, and B. Schoumaker (eds), Théories, paradigmes et courants explicatifs en démographie, Chaire Quetelet 1997, Academia-Bruylant / L'Harmattan, Louvain-la-Neuve, pp. 331-357.

Schoumaker, B. (2001), Une analyse multi-niveaux dynamique de la fécondité légitime au Maroc rural, IUSSP General Population Conference, Salvador (Brazil).

Schoumaker, B. and Tabutin, D. (1999), Analyse multi-niveaux des déterminants de la fécondité. Problématique, modèles et applications au Maroc rural, in La population africaine au 21e siècle, vol. 1, UEPA-NSU (eds), Dakar, pp. 299-332.

Schryock, H.S., Siegel, J.S. et al. (1973), The methods and material of demography, 2 vols, US Bureau of the Census, Washington D.C.

Schultz, T.W. (1973), New economic approaches to fertility: Proceedings of a Conference, June 8-9, 1972, Journal of political economy, 81 (2, Part II).

Schweder, T. (1970), Composable Markov processes, Journal of Applied Probabilities, 7, pp. 400-410.

Singer, J.D. and Willet, J.B. (2003), Applied longitudinal analysis, Oxford University Press, Oxford, New York.

Singleton, M. (1999), Les sens et les non-sens d'un nominalisme démographique, in D. Tabutin, C. Gourbin, G. Masuy-Strobant, and B. Schoumaker (eds.), Théories, paradigmes et courants explicatifs en démographie, Chaire Quetelet 1997, Academia-Bruylant / L'Harmattan, Louvain-la-Neuve, pp. 15-39.

Snijders, T.A. and Bosker, R.J. (1999), Multilevel analysis: an introduction to basic and advanced multilevel modelling, Sage, London.

Steele, F., Diamond, I., and Wang, D. (1996), The determinants of the duration of contraceptive use in China: a multilevel multinomial discrete-hazards modelling approach, Demography, 33, 1, pp. 12-23.

Suppes, P. (1981), Logique du probable, Flammarion, Paris.

Süßmilch, J.P. (1741), Die Göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, Berlin.

Süßmilch, J.P. (1979), "L'Ordre Divin" aux origines de la démographie (1765 ed.), French transl. by M. Kriegel, J. Hecht (ed.), 3 vols, INED, Paris.

Süßmilch, J.P. (1998), L'Ordre Divin dans les changements de l'espèce humaine, démontré par la naissance, la mort et la propagation de celle-ci, annotated French transl. by J.M. Rohrbasser of complete 1741 text, INED, Paris.

Tackács, L. (1964), Processus stochastiques. Problèmes et solutions, Dunod, Paris.

Thygesen, L. (1983), Methodological problems connected with a socio-demographic statistical system based on administrative records, Bulletin de l'IIS, Madrid, 1, pp. 227-242.

Toulemon, L. and Mazuy, M. (2003), Comment prendre en compte l'âge d'arrivée et la durée de séjour en France dans la mesure de la fécondité des immigrants?, Document de Travail 120, INED, Paris.

Tranmer, M. and Steel, D.G. (2001), Ignoring a level in a multilevel model: evidence from UK census data, Environment and Planning A, 33, pp. 941-948.

Tranmer, M., Steel, D.G., and Fieldhouse, E. (2003), Exploring small area population structures with census data, in D. Courgeau (ed.), Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 121-156.

Travers, R. (1969), An introduction to educational research, Macmillan, New York.

Trollegaard, S. (1995), A step-by-step approach for developing integrated local information systems, IIS-ABS conference, Olympia.

Trussell, J. (1992), Introduction, in J. Trussell, R. Hankinson, and J. Tilton (1992), Demographic applications of event history analysis, Clarendon Press, Oxford, pp. 1-7.

Trussell, J., Hankinson, R., and Tilton, J. (1992), Demographic applications of event history analysis, Clarendon Press, Oxford.

Trussell, J. and Richards, T. (1985), Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure, in N. Tuma (ed.), Sociological Methodology, Jossey-Bass, San Francisco (CA), pp. 242-276.

Trussell, J. and Rodriguez, G. (1990), Heterogeneity in demographic research, in J. Adams, D. Lam, A. Hermalin, and P. Smouse (eds), Convergent Questions in Genetics and Demography, Oxford University Press, New York.

Tuma, N.B. and Hannan, M.T. (1984), Social dynamics. Models and methods, Academic Press, Orlando (FL).

Vaida, F. and Xu, R. (2000), Proportional hazard model with random effects, Statistics in Medicine, 19, pp. 3309-3324.

Valade, B. (2001), De l'explication dans les sciences sociales: holisme et individualisme. In J.-M. Berthelot (ed.), Epistémologie des sciences sociales, Presses Universitaires de France (PUF), Paris, pp. 357-405.

Van Imhoff, E. and Post, W. (1997), Méthodes de microsimulation pour des projections de population, in D. Courgeau (ed.), Nouvelles approches méthodologiques en sciences sociales, Population, 52, 4, pp. 889-932. English ed.: (1998), Microsimulation for population projection, in D. Courgeau (ed.), New methodological approaches in the social sciences, Population: an English Selection, 10, 1, pp. 97-138.

Vaupel, J., Manton, K.G., and Stallard, E. (1979), The impact of heterogeneity in individual frailty data on the dynamics of mortality, Demography, 16, 3, pp. 439-454.

Vilquin, É. (1976), La naissance de la démographie, Population et famille, 39, pp. 145-164.

Wachter, K.W., and Lee, R.D. (1989), US births and limit cycle models, Demography, 26, pp. 99-115.

Walliser, B. (2003), Organizational levels and time scales in economics, in D. Courgeau (ed.), Methodology and epistemology of multilevel analysis, Kluwer Academic Publishers, Boston, Dordrecht, London, pp. 157-174.

Willekens, F.J. (1999), The life course: models and analysis, in L. van Vissen and P. Dykstra (eds), Population issues. An interdisciplinary review, Kluwer Academic / Plenum Publishers, New York, pp. 23-52.

Willekens, F.J. (2001), Theoretical and technical orientations towards longitudinal research in the social sciences, Canadian Studies in Population, Special issue on longitudinal methodology, 28, 2, pp. 189-217.

Willekens, F.J. and Rogers, A. (1978), Spatial population analysis: Methods and computer programs, Research Report 78-18, IIASA, Laxenburg.

Wolfinger, R. (1993), Laplace's approximation for nonlinear mixed models, Biometrika, 80, 4, pp. 791-795.

Wong, G. and Mason, W.M. (1985), The hierarchical logistic regression model for multilevel analysis, Journal of the American Statistical Association, 80, 391, pp. 513-524.

Wunsch, G. (1988), Causal theory and causal modelling, Leuven University Press, Leuven.

Wunsch, G. (1994), L'analyse causale en démographie, in R. Franck (ed.), Faut-il chercher aux causes une raison? L'explication causale dans les sciences humaines, pp. 24-40.

Wunsch, G. and Termote, M.G. (1978), Introduction to demographic analysis: Principles and methods, Plenum, New York.

Zelinsky, W. (1971), The hypothesis of the mobility transition, Geographical Review, 61, pp. 219-249.

Author index

Subject index