CHAPTER

24

Individuals and Context in the Multilevel Approach to Behavioral Analysis

DANIEL COURGEAU
Institut national d'études démographiques (INED), Paris, France

Analysis of multistate data aggregated by socio-occupational group, geographic area, administrative unit, and so on, is based on the assumption that the behavior of individuals is homogeneous in each of these subpopulations. If these conditions hold, it is sufficient simply to use the conventional demographic measures and rates and identify the relations that exist between them. For example, a study might relate the emigration rates for the various regions of a country to the unemployment rates, percentages of farmers, and so forth observed in each region. But we must be careful not to infer that these relationships between aggregate-level values are also valid for the corresponding individual-level characteristics. That would be making the error of reasoning known as the *ecological fallacy*—a positive association between a region's emigration rate and the proportion of farmers in its population in fact tells us nothing about the migration probability of farmers, for whom the correlation may be negative. Aggregate-level models can also be constructed in which fertility, mortality and interregional migrations are introduced simultaneously (see Chapter 20 on *r*-level population models), under the additional assumption of independence between the phenomena being studied.

The hypothesis of homogeneity of behavior within each group becomes unnecessary if we work on individual-level data, because an individual's behavior can then be related to his or her personal characteristics (see the previous chapter on event history analysis, and Chapter 131 on causal models in the final volume of this treatise). In this case, the fact that an individual has moved out of a particular region will be related to factors such as being unemployed, being a farmer, and so on. However, the danger here is committing the *atomistic fallacy*, because the analysis ignores the context in which human behavior occurs. In the real world, this context influences individual behavior and it is a mistake to consider individuals in isolation from the constraints imposed by the groups and settings in which they belong.

What is needed, therefore, is an analysis that works at different levels of aggregation simultaneously but whose objective remains that of explaining individual behavior. The pitfall of the ecologic fallacy is thus avoided, because the aggregated characteristic is used to measure a construction that is different from its equivalent at the individual level. It is introduced not as a proxy variable but as a characteristic of the subpopulation that influences the behavior of an individual member of that group. At the same time, the atomistic fallacy ceases to be a problem once the analysis accounts properly for the context in which the individual lives.

I. INDIVIDUAL AND AGGREGATE MEASURES: CONTEXTUAL ANALYSIS

The behavior to be analyzed here is still individual but the explanatory variables can be both individual

and aggregate. Let us begin by looking at the different types of aggregation that can be employed.

1. Types of Measures

For any aggregation level, we can simply add together the individual characteristics and estimate the percentages and averages. Examples are the percentage of farmers, the average starting salary in an occupation, and so on. More complex analytical procedures can also be applied. Thus, as well as average income, the product of this and the binary variable for whether the individual is a farmer could be introduced, which, as will be seen shortly, gives a better characterization of the interaction between individual- and aggregate-level characteristics, or the correlation between this income and household size.

Other characteristics for a given level of aggregation are more general by nature and do not apply at the individual level. Examples are the number of hospital beds in a particular region, or the number of classes in a school. But although these do not correspond to any individual characteristic, they can nonetheless be aggregated at other levels. Thus the number of hospital beds in a region is the sum of the number of beds in each department of the region.

Another class of characteristic is defined for a single level of aggregation and cannot be aggregated at higher levels. The political orientation of a commune, as defined by the party affiliation of its mayor, for example, cannot be aggregated with those of neighboring communes, which may cover a broad spectrum. This characteristic therefore cannot exist at the individual level, because the individual may have voted for another candidate, nor does it exist at the level of the department, because the different communes of the department are not usually all of the same political orientation. Nonetheless, it may influence the behavior of the individuals who live in the commune.

2. A Regression Model

Having specified the characteristics to be evaluated, we will now consider the analytic models that can be used to measure their impact. We begin with an extension of conventional estimation procedures, such as multiple linear regression or logistic regression. This extension is usually termed *contextual analysis*.

This model explains individual behavior by reference to both individual and aggregate characteristics. Let us first of all see how it is formulated in the case of a regression that introduces variables considered at different levels. The example used here is taken from the 1992 Demographic and Health Survey (DHS) of Morocco (Enquête Nationale sur la Population et la Santé, or ENPS-II), for the analysis of fertility determinants in a rural environment (Schoumaker and Tabutin, 1999). Individual fertility is measured by the DRAT (duration ratio), obtained by dividing the number of children each married woman has had by the theoretical number of children she would have had (adjusted for her age and length of union) in a regime of natural fertility. These data are organized in the 72 sampling clusters, which are treated as separate groups.

We write as $y_{i,j}$ the DRAT at the time of the survey of a woman i living in group j. This fertility will first be related to the fact of her husband being a farmer or not ($x_{i,j}$):

$$y_{i,j} = a_0 + a_1 x_{i,j} + e_{i,j} \qquad \text{(Eq. 1)}$$

where $e_{i,j}$ is an individual random term.

The estimated parameter values are reported in the first column of Table 24–1, with their standard error in brackets. According to this conventional regression model, women married to a farmer have a higher fertility than the others. But these parameters may differ depending on the groups and we can estimate them separately for each group j, with the following regression models:

$$y_{i,j} = a_{0j} + a_{1j} x_{i,j} + e_{i,j} \qquad \text{(Eq. 2)}$$

where $e_{i,j}$ is a group random term.

Figure 24–1A displays the results of this estimation. The great diversity in the parameters is clearly apparent, but no conclusion is possible, because most of the parameters a_{1j} are not significantly different from zero. The numbers observed in each group are too small on which to base a conclusion: The differences between the slopes may be the result of purely random error. The solution is to work on the entire population and introduce new characteristics: the percentage of women married to a farmer in cluster $j(x_{.,j})$ and its interaction with the fact of the husband being a farmer, thereby introducing an interaction between an individual and aggregate characteristic. The new model is written

$$y_{i,j} = a_0 + a_1 x_{i,j} + a_2 x_{.,j} + a_3 x_{.,j} x_{i,j} + \varepsilon_{i,j} \qquad \text{(Eq. 3)}$$

where $\varepsilon_{i,j}$ is an error term representing the effects of all the implicit or unobserved variables. This model can of course readily be extended to include additional explanatory variables that are considered to act simultaneously on the female fertility index. Parameter estimation is by the least squares method, with the usual assumptions made for the residuals (i.e., a normal distribution of expectation zero and a constant variance σ^2 for all values of the explanatory variables). Let us

TABLE 24–1 Comparison of results from contextual and multilevel models (rural Morocco, 1992)

Characteristics	Contextual models Estimation (standard error)		Multilevel models Estimation (standard error)	
Fixed				
Constant	0.786 (0.009)	0.749 (0.016)	0.800 (0.014)	0.751 (0.027)
Husband farmer	0.060 (0.013)	0.047 (0.033)	0.040 (0.014)	0.067 (0.036)
% husbands farmer		0.112 (0.040)		0.122 (0.059)
Interaction husband farmer * % husbands farmer		–0.028 (0.061)		–0.072 (0.067)
Random at group level				
σ^2_{u0} (Constant)			0.0084 (0.0024)	0.0076 (0.0024)
σ_{u01} (Covariance)			–0.0017 (0.0020)	–0.0014 (0.0019)
σ^2_{u1} (Husband farmer)			0.0003 (0.0024)	0.0001 (0.0023)
Random at individual level				
σ^2_{e0}	0.1090 (0.0031)	0.1085 (0.0031)	0.1022 (0.0030)	0.1022 (0.0030)
–2 (log-likelihood)	1,530.20	1,519.04	1,455.49	1,451.23

Sources: ENPS-II survey, Morocco, 1992; Schoumaker and Tabutin, 1999.

examine the interpretation of the parameter estimates in more detail.

Calculating the mean of all the observations of group j, gives the following relation, which is valid for each group considered separately:

$$y_{.j} = a_0 + (a_1 + a_2)x_{.j} + a_3x^2_{.j} + \varepsilon_j \qquad \text{(Eq. 4)}$$

where ε_j is the mean error term for each group. We then see that the estimations of the mean values are not distributed randomly but lie along a parabola or a line, if a_3 is zero, when they are plotted as a function of $x_{.j}$. Again for group j, we can now observe the position of the individuals in this group relative to the mean values that have just been calculated. By subtracting relations (3) and (4), we obtain the following expression:

$$y_{i,j} - y_{.j} = (a_1 + a_3x_{.j})(x_{i,j} - x_{.j}) + \varepsilon_{i,j} - \varepsilon_j \qquad \text{(Eq. 5)}$$

which shows that in each group the estimations lie broadly on lines of slope $a_1 + a_3x_{.j}$. We can easily verify that these lines pass through a fixed point if a_3 is not zero or that they are parallel to each other in the opposite case.

For the case examined here, the various parameter estimations are given in the second column of Table 24–1. We see first of all that the lines corresponding to each group are parallel to each other because parameter a_3 is not significantly different from zero: There is no interaction effect between individual fertility and that of the group. On the other hand, fertility is no longer identical regardless of group but is strongly differentiated: As the percentage of farmers among men in the group rises, so fertility increases. However, when this increase is allowed for, the relationship between fertility and the fact of the husband being a farmer is reduced by a quarter (coefficient of 0.047 instead of 0.060). Figure 24–1B displays these lines, which can be compared with the regressions calculated on each group separately (Figure 24–1A). From this we can see that while contextual analysis does take into account the different levels at which the groups are situated, it is less effective at explaining the variation in the slopes corresponding to the various groups.

As it stands, Equation 3 introduces only a single error term at the individual level, which means that the results for individuals in a group are analyzed as if they are independent. In addition, the effects of the aggregate characteristics have a predetermined form—quadratic or linear for the mean, and convergent or parallel lines for the group effect. To dispense with these restrictions, the next step is thus to introduce the random variables specific to each level of aggregation.

II. INTRODUCING GROUP EFFECTS: MULTILEVEL MODELS

It is reasonable to think that the result for an individual in a group may depend on the results obtained by the other individuals in the same group. Because they overlook this intragroup dependence, contextual models produce biased variance estimates for the contextual and individual effects, resulting in confidence intervals that are too narrow. This problem of intragroup dependence can be handled by introducing random effects into the previous contextual models. Let us examine this for the regression discussed in Equation 2.

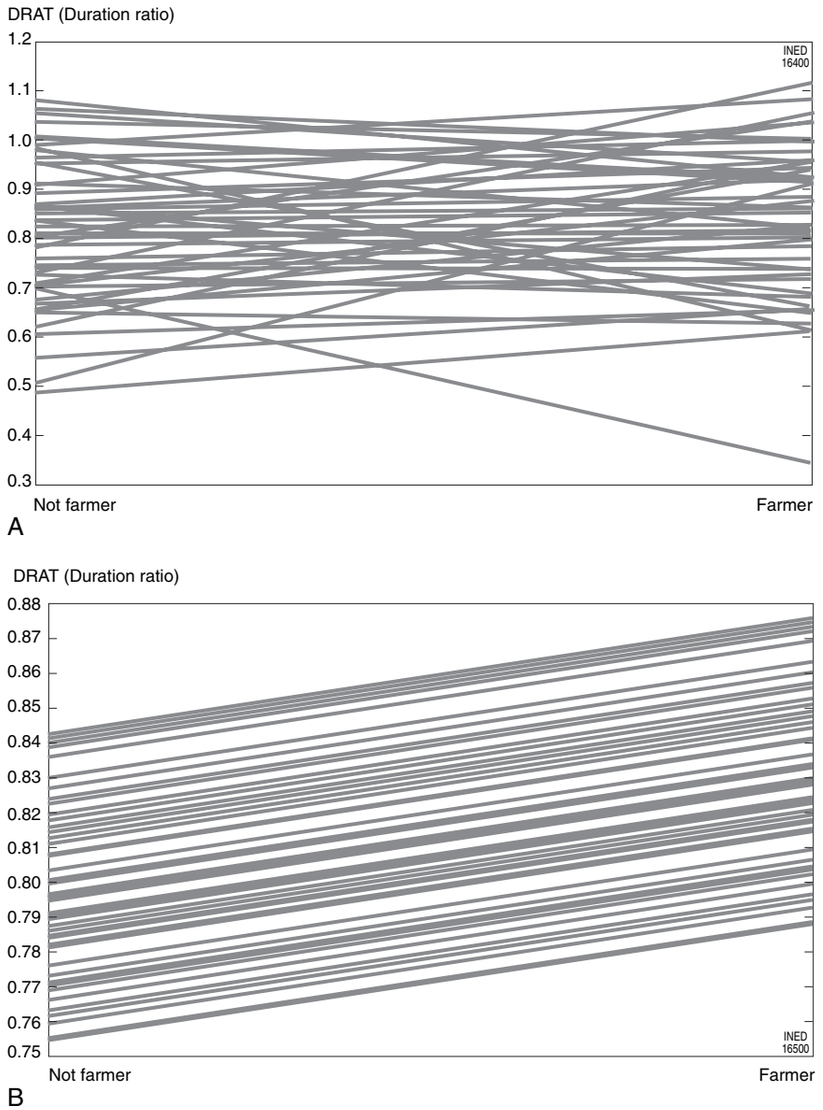

FIGURE 24–1 Fertility estimations by whether the husband is a farmer. **A:** Separate regression on each context. **B:** Contextual regression.

1. Reworking the Previous Regression Model

Because of the small number of individuals observed in each group, the regression parameters estimated earlier will be treated as a random sample from a larger set of parameters. This results in the parameters of Equation 2, a_{0j} and a_{1j}, becoming random variables at the level of the groups, for which we want to estimate the mean a_0 and a_1 and the variation around this mean. We can then write:

$$a_{0j} = a_0 + u_{0j},\ a_{1j} = a_1 + u_{1j} \qquad \text{(Eq. 6)}$$

where u_{0j}, u_{1j} and $e_{i,j}$ are random variables of expectation zero and with the following variances and covariances:

$$\text{var}(u_{0j}) = \sigma^2_{u0},\ \text{var}(u_{1j}) = \sigma^2_{u1},\quad \text{cov}(u_{0j},u_{1j}) = \sigma_{u01} \qquad \text{(Eq. 7)}$$

these three terms corresponding to the level of the groups, and $e_{i,j}$ (Equation 1) to the individual level. This model is fully *multilevel*. The parameters of such a model can be estimated using generalizations of the least squares method (Goldstein, 1995, Courgeau, 2004).

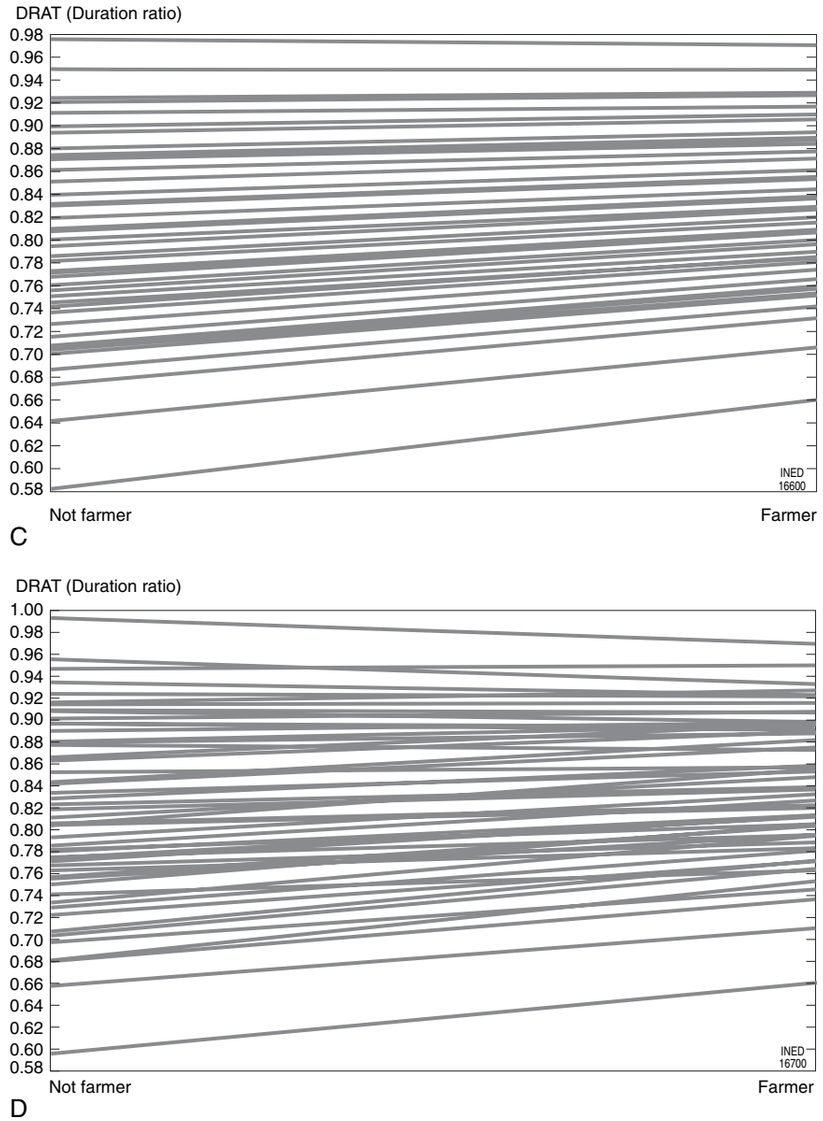

FIGURE 1 (*Continued*) **C:** Multilevel model. **D:** Multilevel model with all variables. Source: Schoumaker Bruno and Tabutin Dominique, 1999. Analyse multi-niveaux des déterminants de la fécondité. Problématique, modèles et applications au Maroc rural, *in:* UEPA-NSU (ed.), *La population africaine au 21e siècle*, vol. 1, p. 299–332. Dakar, UEPA, 630 p. (Proceedings of the Third African Population Conference, Durban).

Applying this model to the previous data gives the results set out in the third column of Table 24–1. The multilevel model, which introduces only the constant and the fact of the husband being a farmer, gives an effect for the latter characteristic very similar to that obtained with the contextual model that introduces all the other characteristics (0.040 compared with 0.047). This is because the introduction of the constant term in the random variables at the level of the groups confirms the effect of these groups on fertility, without the additional terms having to be introduced. Figure 24–1C shows the fertility predicted by this model, depending on whether or not the husband is a farmer. Although the results for the random variable corresponding to having a farmer husband are not significant, the strong negative correlation, close to –1, between the constant and the fact of having a farmer husband, is now reflected in the *fanning-in* configuration of the lines corresponding to each group—The higher the fertility of the nonfarmers, the greater the reduction in the slope of these lines. This result is not obtained with the contextual model, due to its

overrestrictive specification—it is closer this time to the regression lines of Figure 24–1A. Also observed, as expected, is an increase in the standard errors of the fixed parameters, compared with the estimations from Equation 1.

The last column of Table 24–1 introduces the percentage of men who are farmers and the product from multiplying this with the fact that the individual is himself a farmer, again in a multilevel model. The variances of all the random terms are seen to decrease. This confirms that these aggregate characteristics account for part of the random terms at the group level as well as reducing the effect of the farmer husband. However, a significant effect is still associated with the random term corresponding to the constant. Figure 24–1D displays the results predicted by this model, which are very similar to what was obtained in Figure 1A, although some significant parameters from the multilevel model used are now also given.

We have discussed these results at length to demonstrate the utility of multilevel regression. This analysis must be continued with other characteristics, such as educational attainment, standard of living, and so forth of the wife or of her household to see whether this random effect persists and to show the effect of these variables on the fertility level being studied.

2. Analysis for Binary or Polytomous Data

Now let us look at how to construct, for example, a *multilevel logistic model*. For this purpose we use a practical example, drawn this time from the Norwegian population register.[1] We examine interregional migrations over a 2-year period (1980–1981) by women born in 1958 and resident in Norway in 1991.

Let us assume that we are working with the probability that a characteristic to be estimated, $y_{i,j}$, in this case, the fact of having migrated, is 1. Individual i is present in region j before migrating. We want to examine the relationship between this probability and an explanatory variable, $x_{i,j}$, which is assumed here to be binary. This probability, conditioned by the fact of the individual having the characteristic $x_{i,j}$, is written as follows:

$$P(y_{i,j} = 1|x_{i,j}) = p_{i,j} = [1 + \exp(-[a_0 + u_{0j} + (a_1 + u_{1j})x_{i,j}])]^{-1} \quad \text{(Eq. 8)}$$

It follows that the answers $y_{i,j}$, are distributed according to a binomial distribution:

$$y_{i,j} \sim B(p_{i,j},1) \quad \text{(Eq. 9)}$$

[1] We would like to thank the Norwegian statistical services for allowing us to use the data files produced from the Norwegian population registers and censuses.

In this case we have the following conditional variance:

$$\text{var}(y_{i,j}|p_{i,j}) = p_{i,j}(1 - p_{i,j}) \quad \text{(Eq. 10)}$$

The model then becomes a nonlinear model:

$$y_{i,j} = p_{i,j} + e_{i,j}z_{i,j}, \quad \text{where}$$
$$z_{i,j} = \sqrt{p_{i,j}(1-p_{i,j})}, \quad \text{and} \quad \sigma_e^2 = 1 \quad \text{(Eq. 11)}$$

The individual level variance in this case is equal to unity, and we will work principally on the regional level variances and covariances.

In Table 24–2 we have estimated the parameters of the contextual logit and multilevel models, for studying the migration probabilities of the 19 Norwegian regions according to whether the women have at least one child before this migration.

The first contextual model yields a lower migration probability for women with at least one child (first column), close to that obtained for the fixed part of the multilevel model (column 3), although with a smaller variance for the first model. The random part of the multilevel model shows a variation by region and by whether or not the woman has at least one child. We show the joint influence of the fixed and random parameters at the regional level by calculating the logit functions for the probabilities of emigrating from region j, according to whether the woman is childless or not. The logit function of women without children, Π_{0j}, is simply the sum $a_0 + u_{0j}$; its between-region variance is equal to σ^2_{u0}. The logit function of women with children, Π_{1j}, is given by the sum $a_0 + a_1 + u_{0j} + u_{1j}$; so its between-region variance is equal to $\sigma^2_{u0} + 2\sigma_{u01} + \sigma^2_{u1}$.

The data in Table 24–2 can be used to calculate the between-region variance of women with at least one child (0.205), which is four times higher than that of women with no children (0.051). Figure 24–2 displays the values of Π_{0j} and Π_{1j}, corresponding to women with and without children, linked by a line for each region. These lines are characterized by a fan-shaped form, which is explained by the differences between the variances and by the positive correlation between the random variables, corresponding to women with and without children, close to unity (0.93).

When the aggregate characteristics are introduced, these are highly significant in the contextual model (column 2). Figure 24–3 presents the logits of the probability of migrating according to the regional percentage of women with children. This shows that the migration probabilities depend both on whether women have children and on the proportion of women with children living in the different regions. The higher this proportion, the more these probabilities are reduced by the fact of having children. A region's high

TABLE 24–2 Migrations in Norway, whether or not women have children

Characteristics	Contextual models Estimation (standard error)		Multilevel models Estimation (standard error)	
Fixed				
Constant	–1.464 (0.018)	–0.962 (0.062)	–1.503 (0.056)	–1.235 (0.027)
With children	–0.973 (0.048)	–0.125 (0.185)	–0.992 (0.080)	–0.222 (0.349)
Proportion with children		–0.020 (0.002)		–0.009 (0.009)
With children *proportion with children		–0.029 (0.002)		–0.027 (0.013)
Random at regional level				
σ^2_{u0} (constant)			0.051 (0.029)	0.049 (0.022)
σ_{u01} (covariance)			0.042 (0.027)	0.032 (0.015)
σ^2_{u1} (with children)			0.070 (0.068)	0.038 (0.029)
–2 (log-likelihood)	22,934	22,808	19,237	19,237

Source: Norwegian population register, Central Bureau of Statistics, Oslo.

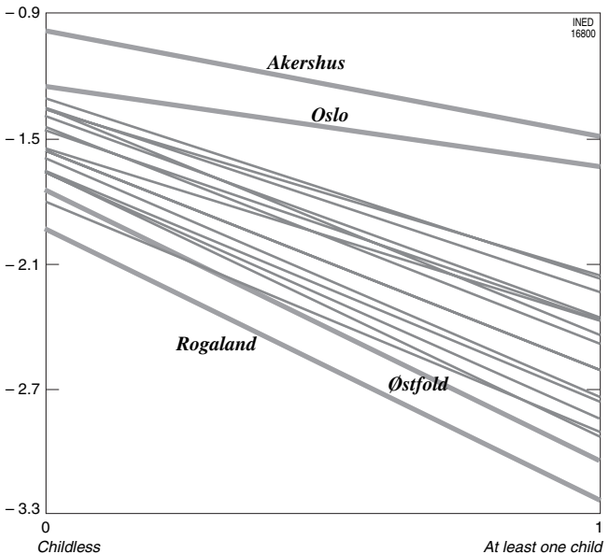

FIGURE 24–2 Logit model on Norwegian migrations: logit of the probability of migrating by number of children.

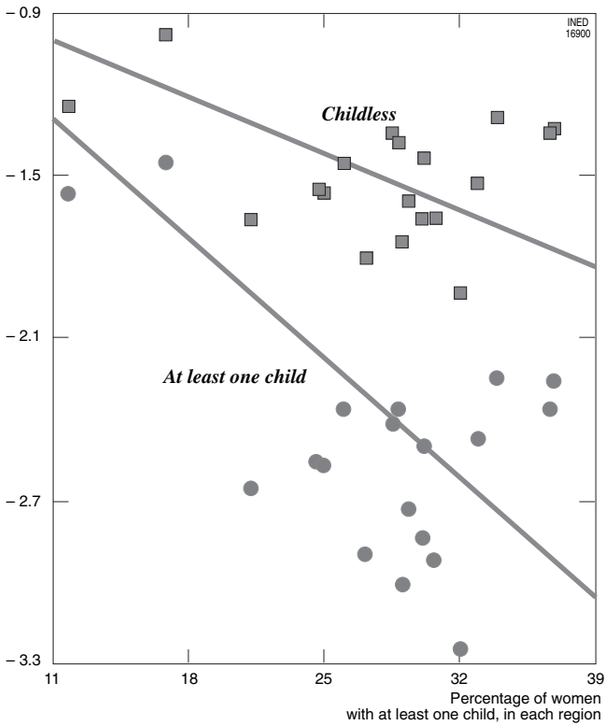

FIGURE 24–3 Logit model on Norwegian migrations: logit of the probability of migrating, relative to the percentage of women with at least one child.

aggregate fertility is associated with a greater capacity to retain women with children.

These conclusions are completed by the multilevel model. Only the interaction effect between the individual and aggregate characteristic is significant on the fixed parameters, indicating as before that the higher the proportion of women with children in a region, the less likely these women are to migrate. At the same time, a reduction in variance is observed at the regional level for women with children (0.151 compared with 0.205), which confirms the effect of the aggregate variable. However, the regional effect for the constant is still significant and at the same level (0.05): this was not properly accounted for in the contextual model. Figure 24–3, which also displays the logits of the probability of migrating for each region as a function of the percentage of women with children, reveals a more complex situation than that obtained with the contextual logit regression: the points corresponding to the different regions are highly dispersed in relation to the regression lines plotted on the same diagram.

Various characteristics (marital status, labor market status, educational attainment, etc.) can of course be

introduced, separately or simultaneously, to examine their impact on the probability of interregional migration. Their effects may differ from those we have identified by introducing the presence of children. Thus for men the fact of working in farming greatly reduces the probability of emigrating (Courgeau and Baccaïni, 1997). Conversely, it has no effect on the random variables at the regional level: In this case the lines, corresponding to those shown on Figure 24–2, will be parallel to each other. Introducing the percentage of farmers present in each region now produces two parallel lines that rise with this percentage: farmers have a much lower probability of migrating than other categories, but the higher the percentage of farmers in a particular region, the higher the probability of migrating for all categories. This result clearly illustrates the danger of inferring results for individuals from results obtained at a more aggregated level. Having a large number of farmers in a region results in a higher probability of migrating for all population categories, doubtless due to the increased scarcity of nonagricultural employment in such regions. However, this does not mean that farmers have a higher probability of emigrating than the other categories. Indeed, the opposite is observed.

Finally, such analyses can be extended to competing risk models in which individuals experience several types of events, such as death from a specific cause, emigration to different regions, and so forth. These models are a relatively straightforward generalization of the binomial model presented above and we will not discuss them in detail here.

3. Multilevel Event History Analysis

We now come to the most comprehensive analysis of behavior, which combines an event history approach (see Chapter 23 on *event history analysis*) and a multilevel analysis. Because individuals are followed throughout their occupancy of a given state, some of their own characteristics may change at certain points in time (they marry, change occupation, etc.) and the characteristics of the regions in which they live will be subject to continuous change over time (increase in the percentage of married people, changes in the percentages of occupational groups, etc.).

Conducting such an analysis calls for finely detailed survey data, of the kind obtained from retrospective designs, that record all the events in the life course of individuals. Population register data, even when linked to census data as in the Norwegian source used in the prior section, can be used to monitor only a small number of events. It is inadequate to sustain a full multilevel event history analysis. Also, for event history surveys to be usable they must be on large numbers. The data must be capable of revealing the differences between a large number of groups, geographic regions, and so forth.

To illustrate an analysis on these lines we will use data from the Youth and Careers survey conducted by INSEE on a sample of nearly 20,000 individuals. The event we examine is leaving home for the first time by young girls (9043). France has been divided into 19 regions.

The exposition uses a semiparametric model, also known as the Cox model. The objective of this model is to estimate an instantaneous hazard rate (in this case for leaving the parental home), at $t_{i,j}$, as a function of the column vector of the different characteristics $Z_{i,j}$ and the regions j in which the individual i lives. This hazard rate, $h(t_{i,j};Z_{i,j})$, can be written:

$$h(t_{i,j};Z_{i,j}) = h_0(t_{i,j})\exp(\beta_j Z_{i,j}) \qquad \text{(Eq. 12)}$$

In this case, some of the parameters for estimation of the row vector, β_j, will have a random part.

Let us suppose that the times at which a young woman leaves the parental home, or is lost to observation without yet having left, are time-ordered and that at each point in time we can determine the population exposed to the risk. At time $t_{i,j} = l$, the probability that a young woman i will leave her parents, conditioned by the population exposed to the risk and by a departure from the parental home occurring at this date, is equal to:

$$\frac{h_0(l)\exp(\beta_j Z_{i,j})}{\sum_{k\in R_i} h_0(l)\exp(\beta_j Z_{k,j})} \qquad (13)$$

where R_i is the complete population of individuals exposed to the risk at $t_{i,j} - 0$ in any group.

Cancelling out $h_0(l)$, which is in both numerator and denominator, we obtain a partial likelihood that no longer depends on this baseline hazard by multiplying these conditional probabilities at all the dates. What we have written here is a *multilevel event history model*.

So far we have assumed that a single event is observed at each point in time. In practice data tend to be more grouped and we observe $n_{i,j}$ departures from the parental home at date l. The parameters β_j can still be estimated by the maximum likelihood method. The values of $h_0(l)$ are then estimated using these values of $\hat{\beta}_j$. Estimation of the latter function can be simplified by fitting a polynomial distribution (Goldstein, 1995), thus giving in effect a parametric model. A third-degree function is used here.

The explanatory variables can be defined at any level of aggregation and can be considered time

dependent. Those introduced here are the individual's birth cohort, whose effect is represented by a second-degree function; the number of brothers and sisters, up to a maximum of four; and the fact of the young woman being in employment before leaving the parental home, which is a time-dependent variable.

Table 24–3 presents the estimated parameter values depending on whether a simple or multilevel event history model is used. The fixed parameters are very similar in both cases. They indicate a slight fall in age at leaving the parental home, up to the cohorts born at the end of the 1950s, followed by a sharp increase, in parabolic form, of this age for subsequent cohorts. The more brothers and sisters there are, the earlier leaving the parental home occurs. Last, for the woman in employment, the probability of leaving the parental home is multiplied by 1.9 (exp [0.64]). The multilevel model shows that depending on the individual's region of residence there is substantial variation in the probabilities. Figure 24–4 shows how this multiplicative effect varies from region to region. The highly urbanized regions are those where leaving the parental home occurs latest (regions around Bordeaux, Marseille, Paris), whereas the opposite is observed in the rural regions (such as Brittany and Normandy). The cohort effect at the level of the random variables is not significant: The curves on Figure 24–4 are nearly all parallel to each other, even though they were estimated with this random effect.

The example given above is of course merely the initial stage of such an analysis. Many other characteristics will have an influence on the behavior of young people, including family and parental background, the economic, political and social conditions at the time of taking the decision, and a range of individual characteristics.

TABLE 24–3 Leaving parental home, France, by various characteristics

Characteristics	Simple event history model Estimation (standard error)	Multilevel event history model Estimation (standard error)
Fixed		
Cohort*	–0.109 (0.018)	–0.121 (0.022)
Cohort^2	–0.222 (0.028)	–0.234 (0.017)
Number of brothers and sisters	0.039 (0.008)	0.037 (0.007)
Working prior to departure	0.641 (0.026)	0.635 (0.039)
Random		
σ^2_{u0} (Constant)		0.012 (0.005)
σ_{u01} (Covariance)		0.006 (0.004)
σ^2_{u1} (Cohort)		0.005 (0.004)
–2 (log-likelihood)	42,797	22,684

*Cohort centred on 1964 and divided by 10.
Source: Youth and Careers survey, INSEE.

III. GENERALIZATION OF THE ANALYSIS

In the present exposition, we have deliberately restricted attention to relatively simple models, with the aim of illustrating the contribution that contextual and multilevel approaches can make to an exclusively individual approach. We will now indicate the extensions that can be given to these models.

So far we have considered hierarchical aggregation at a maximum of two levels: the individual and rural clusters; the individual and geographic regions. If we retain this *hierarchical classification* we can begin by considering a greater number of levels of aggregation. Thus the individual is a member of a household, which is resident in a neighborhood, which in turn is located in a town, and so forth. However, any other structure of classification can also be used. For example, pupils are grouped by classes, the classes are in schools, the schools may be public or private, and so forth. The

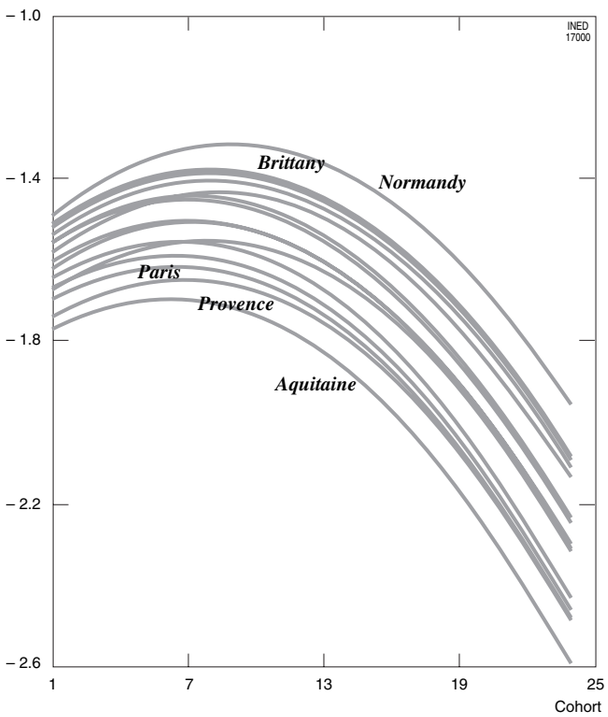

FIGURE 24–4 Logarithm of the probability of migrating by Region.

explanatory characteristics, individual or aggregate, will operate at these different levels, with the random variables specific to each. In each case, careful thought is required about how to interpret the action of a characteristic that operates at different levels of aggregation.

More complicated structures of classification can be realized, such as individuals resident in towns that are grouped by ascending size order but also by their role as centres of government, manufacturing, tourism, etc. The result here is a *cross-classification* in which towns are classified both by their size and by their function. The models presented earlier can be generalized to classifications like these (Goldstein, 1995; Courgeau, 2003).

It is of course possible to have data structures that are in part hierarchical and in part cross classified. For example, individuals can be classified by type of residential area (town centre, suburb, etc.) and type of workplace (industrial, commercial, etc.), which are themselves units in a hierarchical classification of departments and regions.

We can also try to dispense with the hypothesis that underlies the previous analyses, namely, that the groups have no structure. This hypothesis is untenable when working on small groups such as the family and the household. In this case we need to introduce a social structure for these groups, by distinguishing the behavior of parents from that of children and of other people living in the group, for example (Lelièvre *et al*., 1997).

Looking still further ahead, there is also a need to advance beyond the individual approach used here, which explains behavior by characteristics measured at different levels of aggregation. The analysis must be extended to examine the elaboration and modes of operation of the various levels. How are these levels organized, and how is their action modified? For example, actions by isolated individuals in a given community may produce an awareness of problems and action to resolve them by those with influence at the community-wide level (new laws, acceptance of new forms of behavior, and so on). This could create new problems at both the societal and individual levels, for example.

In summary, multilevel analysis offers a solution to many problems encountered when demographic analysis is conducted at either the individual or a more aggregated level. It combines the different levels of aggregation in an entity that is more informative than when each level is considered separately. As such it enables us to assess the conjoint influences of individual and aggregated characteristics on demographic behavior. Further advances require its generalization to the study of the operation of the different levels of aggregation, thereby offering new insights into the nature of change in human societies in their full complexity.

References

Bressoux Pascal, Coustière Paul and Leroy-Audouin Christine, 1997. Les modèles multiniveau dans l'analyse écologique: le cas de la recherche en éducation, *Revue Française de Sociologie*, vol. 38(1), p. 67–96.

Courgeau Daniel and Baccaïni Brigitte, 1997. Analyse multiniveaux en sciences sociales, *Population*, vol. 52(4), p. 831–864. (Also in English: Multilevel analysis in social sciences, *Population: An English Selection*, 1998, vol. 10(1), p. 39–71).

Courgeau Daniel, 1994. Du groupe à l'individu: l'exemple des comportements migratoires, *Population*, vol. 1, p. 7–26. (Also in English: From the group to the individual: what can be learned from migratory behaviour, *Population: An English Selection*, 1995, vol. 7(1), p. 145–162).

Courgeau Daniel, 1999. De l'intérêt des analyses multi-niveaux pour l'explication en démographie, *in:* Dominique Tabutin, Catherine Gourbin, Godelieve Masuy-Stroobant and Bruno Schoumaker (eds.), *Théories, paradigmes et courants explicatifs en démographie*, p. 93–116. Louvain-la-Neuve, Academia-Bruylant/L'Harmattan, 670 p. (Chaire Quetelet, 1997).

Courgeau Daniel, (ed.), 2003. *Methodology and epistemology of multilevel analysis*, Methodos Series, no. 2. Dordrecht/Boston/London, Kluwer Academic Publishers, 236 p.

Courgeau Daniel, 2004. *Du groupe à l'individu: synthèse multiniveau*, Paris, coll. de l'INED, PUF, 242 p.

Diez-Roux Ana, 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis, *American Journal of Public Health*, vol. 88(2), p. 216–222.

Entwistle Barbara and Mason William M., 1985. Multilevel effects of socio-economic development and family planning programs on children ever born, *American Journal of Sociology*, no. 91, p. 616–649.

Goldstein Harvey, 1995. *Multilevel Statistical models.* London, Edward Arnold, 178 p.

Lelièvre Éva, Bonvalet Catherine and Bry Xavier, 1997. Analyse biographique des groupes: les avancées d'une recherche en cours, *in:* Daniel Courgeau (ed.), Nouvelles approches méthodologiques en démographie, *Population*, vol. 52(4), p. 803–830 (Also in English: Event history analysis of groups. The findings of a on-going research project, *Population: an English Selection*, 1998, vol. 10(1), p. 11–37).

Loriaux Michel, 1989. L'analyse contextuelle: renouveau théorique ou impasse méthodologique, *in:* Josiane Duchêne, Guillaume Wunsch and Éric Vilquin (eds.), *L'explication en sciences sociales: la recherche des causes en démographie*, p. 333–368. Louvain-la-Neuve, Éditions Ciaco, 476 p.

Schoumaker Bruno and Tabutin Dominique, 1999. Analyse multiniveaux des déterminants de la fécondité. Problématique, modèles et applications au Maroc rural, *in:* UEPA-NSU (ed.), *La population africaine au 21e siècle*, vol. 1, p. 299–332. Dakar, UEPA, 630 p. (Proceedings of the Third African Population Conference, Durban).

Von Korff Michael, Koepsell Thomas, Curry Susan and Diehr Paula, 1992. Multilevel analysis in epidemiologic research on health behaviors and outcomes, *American Journal of Epidemiology*, vol. 135, p. 1077–1082.