9

HETEROGENEITY AND INTERDEPENDENCE: A TEST USING SURVIVAL MODELS

Christopher Winship

NORTHWESTERN UNIVERSITY AND ECONOMICS RESEARCH CENTER, NORC

The idea that the timing of life-cycle transitions is interdependent is ubiquitous in the life course literature. Elder (1978, p. 26) has gone so far as to define the life course as "a concept of interdependent life patterns which vary in synchronization." The interdependence of life-cycle transi-

This research was supported by grants from the National Science Foundation and the National Institute on Aging. I am particularly grateful to Ian Domowitz, who provided many important and insightful comments on an earlier draft. Many others provided useful comments: James Coleman, Robert Mare, Lars Muus, Larry Radbill, Nancy Tuma, and an anonymous reviewer. Jean Mitchell and Nancy Winship helped with the editing. Any mistakes that remain are of course my responsibility.

tions has been used to explain the age pattern of behavior (Mare, Winship, and Kubitschek 1984; Wohlwill 1970) and has also been a key component of explanations of historical change in family and life-cycle behavior. (See Modell and Hareven 1978 for a review.)

Despite the prevalence of the idea that transitions are interdependent, little theoretical or methodological work has been done to examine either what it means for two transitions to be interdependent or how one might tell whether two transitions are interdependent. My purpose here is to ameliorate this situation by developing a test of whether the timing of two transitions is in fact causally related. This test is developed by modeling interdependence in terms of hazard rates and using methods from the literature on survival analysis and event-history analysis (for example, Kalbfleish and Prentice 1980; Tuma and Hannan 1984).

This chapter contrasts two different explanations of why transitions may appear to be related. First, there may in fact be a causal or structural relation between transitions. With respect to the age of leaving school and age of marriage, for instance, one might hypothesize that nonstudents have better marital prospects than students. As a result, individuals might delay marriage until after they have finished school. The other possibility is that there is in fact no real causal or structural relationship between events. Instead, one observes a relationship between two transitions because individuals differ in the rate at which they approach adulthood. Some individuals may decide to enter adulthood sooner or, to use a psychological term, may mature faster. As such, they are more likely to leave school and marry earlier. For others the reverse may be true. Heterogeneity across individuals in their rates of maturation can induce observed dependence across individuals between the two transitions.

An example will help clarify the distinction between the two types of explanation. The example is of physiological growth. Consider the relationship across individuals between the age at which their heads reach full adult size and the age at which their legs reach full adult length. If we examined these two variables, we would probably find that they are positively correlated. Moreover, we might find that the age at which individuals' heads reach full development is generally earlier than the age at which their legs reach full length. Few nonexperts (and perhaps experts) would argue that there is a direct causal relationship between these two variables. Most people, myself included, would probably argue that individuals differ in their rate of physiological growth and that it is because of individual heterogeneity in growth rates that we observe a relationship between the age at which individuals' heads reach full size and the age at which their legs reach full length.

Now consider the relationship between the age at which individuals reach full adult height and the age at which they reach normal adult weight. Here it is not clear that a causal relationship is absent. Since attaining full height is closely related to development of bone structure, and bone structure is an important element in supporting people's full weight, it may make sense to argue that there is a causal relationship between the two events.

The issue is analogous to the problem of spurious correlation: Are two variables related because there is a direct causal link between them, or are they related as a result of being dependent on a third variable that has not been controlled? In this chapter the point is not, however, that there is a third variable but that there may be a process (unobserved) producing the observed relation between the timing of two transitions.

The idea that some process is creating the observed relationship between the timing of two events is critical to this chapter. For the lack of a more general term I shall refer to this process as maturation. In using this term I intend to connote nothing more than the idea that there is some underlying process, varying across individuals and related to age, that determines behavior. I shall refer to this process as individual heterogeneity in rate of maturation or, for short, individual heterogeneity.

The argument that two variables are related because they are causally linked is common in sociology. It is at the heart of path analysis and the use of simultaneous equations. It is also found in the literature on Markov chains (for example, Coleman 1964; Tuma and Hannan 1984). The idea that there is some underlying process creating observed relations between variables is much less common. In other disciplines, however, it is an important hypothesis. Physiological growth is one example. Psychological and cognitive development theories are another. Economic theories of individual choice over the life cycle represent in disguised form a third example: Behaviors are related because they are part of a single optimization process.[1]

[1] The result presented in this chapter parallels one found in economics in an interesting way. The theories of the consumer and the firm imply an important symmetry condition known as Slutsky symmetry. Slutsky symmetry states that in a maximization model the cross-price elasticities must be equal. The result to be presented here is a symmetry condition, as well. For further discussion of Slutsky symmetry see Varian (1984).

Although the idea that there are underlying age-related processes determining individual behavior is not common in sociology, there is work that proposes a closely related set of ideas. A number of researchers working in the life-cycle tradition have argued that certain behavior is age-graded. The term *age-graded* refers to the descriptive aspects of behavior—that certain behavior varies by age, often dramatically so. Within this literature, though, there is also the suggestion that some behavior is age-determined or age-dependent. One example is the proposition that age norms are important determinants of behavior (Neugarten, Moore, and Lowe 1965; Elder 1975). Another example is the argument of Riley, Johnson, and Foner (1972) that roles are highly stratified by age and therefore age is a critical variable for understanding variations in individual behavior.

There is an important difference between the idea that behavior is partially age-determined and the idea that transitions are embedded in a singular process of maturation. The notion of age-gradedness assumes that individuals of the same chronological age are in similar positions. The idea of maturation allows for the possibility that individuals may change at different rates. Hence individuals of the same chronological age may be at very different positions with respect to their life-cycle development. The concept of maturation also suggests that behavior is age-determined, but in such a way that individuals may have different timetables.

The rest of this chapter develops a set of mathematical models and then derives some results. The next section provides a short introduction to the idea of a hazard rate; causality and heterogeneity are then modeled in terms of hazard functions. I then look at the problem of distinguishing between the causal dependence and individual heterogeneity in rates of maturation.[2] I show that in the most general case it is impossible to

[2] The work in this chapter is related to other work in sociology and econometrics. Tuma (1980), in a Markov chain context, has examined the implications of estimating the parameters of one process when it is related to another process that is not taken into account. In a sense the present discussion is an extension of Tuma's work: I wish to look at the implications of examining the relationship between two processes (transitions) when they are possibly related to a third. The ideas presented here also parallel work on the distinction between heterogeneity and state dependence (Heckman 1978, 1981; Heckman and Borjas 1980). These models consider transitions of a single type that are repeatable. The concern is whether there is any structural relationship across time between events or whether all the observed dependence is simply due to unobserved heterogeneity across individuals. I am also concerned with whether the observed relation between variables is simply a result of heterogeneity. I, however, assume that events are not

distinguish between these two explanations. It is, however, possible to make a distinction if restrictions are made. I develop a test for true dependence and present an empirical example based on the age of leaving school and age of marriage using the Occupational Change in a Generation II (OCGII) Survey.

MODEL SPECIFICATION

Hazard Rates and Hazard Functions. There are a number of different ways to model interdependence and heterogeneity. In the last five years or so, sociologists have come to appreciate the virtues of formulating continuous-time models in terms of hazard functions. These virtues include the fact that problems of right censoring (that for some individuals the event has yet to take place) are easily dealt with and that the effects of time-varying variables can be incorporated.

We begin with hazard rates and functions. (Readers familiar with these ideas may wish to skip this section. The following two sections show how interdependence and heterogeneity can be modeled with hazard functions.) Formally, the hazard rate for an event is defined as

$$h(t) = \lim_{\Delta T \to 0} \frac{P(t \leq T + \Delta T | t \geq T)}{\Delta T} \qquad t \geq 0$$

where the numerator is the probability that the time a transition takes place, t, given it has not occurred before time T, is between T and $T + \Delta T$ (where Δt is an arbitrarily small increment of time). Note that the hazard rate is a function of t. As such, the hazard rate here is a hazard function. Since I am concerned with the timing of events in individuals' lives, time here and throughout the rest of the chapter is equivalent to age. Thus, by letting the hazard be a function of t, the hazard may vary with age.

The hazard rate is equal to the conditional likelihood of an event occurring at t for those individuals for whom the event has yet to occur. This is seen by noting that the hazard is equal to

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{1}$$

---

repeatable. The chapter also relates to recent work on the nonparametric specification of heterogeneity in continuous-time models as random effects (see Heckman and Singer 1984). The results of this work are not applicable to the questions posed here, however.

where the numerator is the density function or likelihood for the variable t and the denominator is 1 minus the cumulative distribution function, $F(t)$. The denominator, as we shall see, is also equal to the survivor function. One can solve for $f(t)$ as follows:

$$f(t) = h(t)\exp\left[-\int_0^t h(z)\,dz\right] \tag{2}$$

Equation (1) shows that one can define the hazard function for any variable if its density function is known (and thus its cumulative distribution function).[3] Equation (2) shows that if one knows the hazard function for a variable its density can also be derived.

One can think of the hazard rate as being the "speed" at which an individual is approaching a particular transition. A hazard rate captures the idea that individuals may differ in their rates of maturation: Individuals who have large hazard rates are likely to undergo transitions earlier; those whose hazards are closer to zero are likely to undergo transitions later.

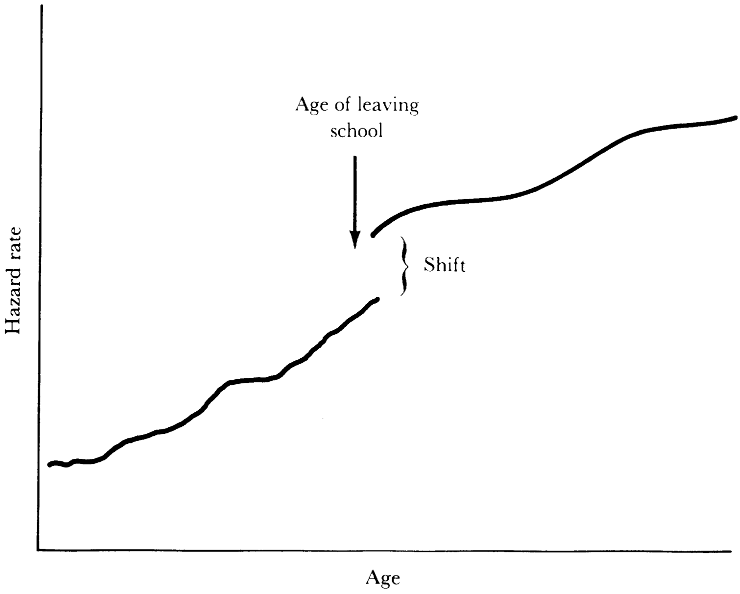
Interdependence. My concern here is to determine whether two transitions are structurally interdependent. Consider two transitions a and b. Generally I want to think of a as the transition of leaving school and b as the transition into an individual's first marriage. Let t_a and t_b be the age at which these two events respectively occur. Besides using a and b as labels I also want to think of them as indicator variables with $a = 1$ and $b = 1$ indicating that the respective event has occurred and $a = 0$ and $b = 0$ indicating that the event has not occurred. The interpretation will be obvious from the context.

Dependence of events on each other can be modeled by assuming that the hazard for each event depends on the time at which the other event occurs. That is: $h_a = h_a(t, t_b)$ and $h_b = h_b(t, t_a)$. This specification is quite general. It allows the occurrence of event b at t_b to effect the hazard for a at time t if t_b is less than or greater than t. Since the hazard for event a depends jointly on t and t_b, any type of functional dependence is allowed. Thus a full set of lead and lag effects can be incorporated.[4]

[3] Throughout the chapter I assume that variables have well-defined densities except where noted otherwise.

[4] It can be shown that any type of dependence can be specified in this way. This can be done by decomposing any joint distribution function into the product of a conditional distribution times a marginal distribution and then deriving how these two terms relate to the two hazards specified above.

FIGURE 1. Graph of hazard for marriage for a hypothetical individual.



A more restricted specification of dependence is to assume that the hazard for one event is simply affected by whether the other event has or has not occurred. This notion can be modeled as $h_a = h_a(t, b)$ and $h_b = h_b(t, a)$, where within each function a and b are being used as indicator variables. This specification captures the idea that when one event occurs it shifts the hazard function for the other event. For example, leaving school might shift the marital hazard function. This is illustrated in Figure 1.

This last notion of dependence is equivalent to that found in the literature on Markov chains where states have been cross-classified by each other (see Coleman 1964; Tuma and Hannan 1984). In terms of the substantive example used in this chapter there are four states: in school/not married, out of school/not married, in school/married, and out of school/married. The interest then is in examining transition rates between these states. If one excludes transitions that involve more than one status change (say, moving from the in school/not married state to the out of school/married state), then the transition from being not married to married involves either moving from the in school/not married state to the in school/married state or from the out of school/not married state to the out of school/married state. If the hazard for these two transitions differs, one might infer that leaving school has increased the marital hazard.

An implication of the concept of dependence found in the Markov chain literature or the more general type specified here is that the timing of the two transitions for the individual is not independent in the statistical sense. That is, if f_a and f_b are the density functions of transitions a and b respectively and f_{ab} is the joint density function, then

$$f_{ab}(t_a, t_b) = f_a(t_a) f_b(t_b) \tag{3}$$

Conceptually, if an individual were to live a number of times, then across those lifetimes t_a and t_b would not be statistically independent.

Lack of statistical independence is a general way of defining interdependence between two events at the individual level and is equivalent to that given above for hazards. I shall say that two events are interdependent (or equivalently that they are structurally related) if, at the individual level, they are not statistically independent—that is, Equation (3) holds.

If one could observe individuals over repeated lifetimes or if there were some analogous set of repeated trials, standard statistical techniques could be used to test whether two events were independent of each other. In fact, it would be possible to consistently estimate various parameters in the presence of unobserved differences between individuals in their hazard functions. (See, for example, Kalbfleisch and Prentice, 1980; Yamaguchi, 1983.) In most social science applications, however, data on repeated trials do not (and often cannot) exist. As a result, the analysis must be done on a sample of individuals. Differences across individuals in their hazard functions then become a potential problem.

Heterogeneity. In the introduction I argued that individuals might differ in how fast they approach adulthood or equivalently in how fast they "mature." I now want to specify this idea in terms of the hazard functions for our two transitions. To make it clear that functions differ across individuals, let i be a *continuous* variable indexing individuals of different types. Individuals with the same hazard functions are considered to be of the same type. Let $K(i)$ be the density function for i (it will be used later in this chapter).[5] The most general specification of the concept

[5] Note that i can be reduced to a discrete index by allowing its density $K(i)$ to have mass only at a finite set of points.

that individuals differ in their rates of maturation would be to assume that the two hazards take the form

$$h_a(i,t) = h(u_a(t), r(i,t))$$
$$h_b(i,t) = h(u_b(t), r(i,t))$$

where u_a and u_b represent components of the hazard common across individuals and $r(i,t)$ represents individual heterogeneity that shifts the hazard through the function h. Since $r(i,t)$ is solely a function of i and t, it can be interpreted as an individual-specific age effect on the hazard. It captures the premise that individuals at different ages differ in their rates of maturation.

As stated the specification is quite general. It is worth considering several examples that are common in the literature on hazard functions. These are the additive hazard, the multiplicative or proportional hazard, and the accelerated failure-time model:

$$\left.\begin{aligned} h_a(i,t) &= u_a(t) + r(i,t) \\ h_b(i,t) &= u_b(t) + r(i,t) \end{aligned}\right\} \quad \text{(additive hazards)}$$

$$\left.\begin{aligned} h_a(i,t) &= u_a(t) r(i,t) \\ h_b(i,t) &= u_b(t) r(i,t) \end{aligned}\right\} \quad \text{(multiplicative hazards)}$$

$$\left.\begin{aligned} h_a(i,t) &= u_a(r(i,t)t) \\ h_b(i,t) &= u_b(r(i,t)t) \end{aligned}\right\} \quad \text{(accelerated failure-time model)}$$

In the additive hazard formulation, differences in maturation rates affect the two hazards additively. Since $r(i,t)$ is a function of t, the difference between the two hazards can change with age. Across individuals, however, it is assumed that at any time the difference between the two hazards is the same. Thus it is assumed that the hazards for the two events are the same across individuals except for the addition of a term $r(i,t)$ representing differences across individuals in their rates of maturation. Instead of assuming that differences between individuals enter additively, the multiplicative hazard model assumes that they enter multiplicatively. The ratio of hazards across individuals is then constant.

Both the additive and multiplicative models operate directly on the hazard. The accelerated failure-time model assumes that heterogeneity enters as a direct transformation of time. The simplest form for $r(i,t)$ is as a constant, $r(i)$. In this case, $h_a(i,t) = u_a(r(i)t)$ and $h_b(i,t) = u_b(r(i)t)$. Here $r(i)$ simply stretches or shrinks the time scale by a multiplicative factor. Note that the accelerated failure-time model allows for any transformation of t since I can let $r(i, t) = r^*(i, t)/t$, which gives $h_a(i, t) = u_a(r^*(i, t))$ and $h_b(i, t) = u_b(r^*(i, t))$. In general $r^*(i, t)$ is restricted to be a strictly increasing monotonic function. As with the additive and multiplicative models, the accelerated failure-time model assumes that the hazard for the two events is transformed in the same way.

In all three models I have assumed that, net of the effect of $r(i, t)$, the hazards for the two events are the same for all individuals. This implies that the "relation" between the hazards for the two individuals is the same across individuals. This is a strong and probably unrealistic assumption. Across groups, transitions have different relative positions in the life cycle. For instance, cohorts born early in this and the last century tended to leave school relatively early and marry relatively late. More recent cohorts, however, have tended to leave school later and marry earlier (Hogan, 1981). This type of change cannot be explained by differences across cohorts in the rate at which they approach adulthood. Rather, the transition out of school and the transition into marriage must have a different relationship across cohorts.

One approach to solving this problem would be to let the relationship between the two transitions be a function of observed variables X. For instance, one might want to condition the relation between the age of leaving school and age of marriage on cohort, ethnicity, region, or perhaps other variables as well. I have not done this in the analyses presented later in the chapter, but the appendix outlines how it could be done.

MODEL IDENTIFICATION

Specification. The concept of structural dependence as discussed in the last section can be combined with the idea that there are individual differences in rates of maturation into a single set of equations where the u_a and u_b terms are used to capture differences between h_a and h_b:

$$h_a(i, t) = h(u_a(t), r(i, t), t_b)$$

$$h_b(i, t) = h(u_b(t), r(i, t), t_a)$$

The basic issue is whether one can distinguish between the effects of the different components in this model. In particular, is it possible to determine whether t_a and t_b affect each other or whether all the observed dependence between t_a and t_b is the result of $r(i, t)$? In asking this question I assume that transitions are not repeatable and that the possible effects of exogenous variables are not being considered.[6]

There are two specific questions one might ask in attempting to answer the more general question. First, when there is heterogeneity, but no structural dependence between the timing of two transitions at the individual level, are there any restrictions on the observed relationship between the timing of the two transitions in the population? If the answer to this question is yes, then it may be possible to develop a test of the null hypothesis that the observed relationship between two events is the result simply of heterogeneity. Second, if this is the case, how easily can this possibility be distinguished from the alternative that there is interdependence between the two events at the individual level? In the language of statistics: What is the potential power for a test to distinguish the null hypothesis from the alternative of interdependence?

The next section presents an example that illustrates the problem. The following two sections examine whether heterogeneity with no interdependence imposes any restrictions on the observed relationship between two events. In the case where no restrictions are placed on how heterogeneity enters the hazard, the null hypothesis puts no restriction on the observed relationship between two events. This is a very important finding —it means that in general one cannot distinguish between heterogeneity and true interdependence. In the following sections, I examine the implications of additive heterogeneity. Here the observed relationship between two events is restricted. This then allows one to develop a test of the null hypothesis of no interdependence at the individual level. Although the power of this test is not formally analyzed, in empirical analyses in which the alternative appears to be well specified (for example, that leaving school increases the marital hazard), the test appears to have no problem distinguishing between the null and the alternative hypotheses.

An Example. To appreciate this problem consider the preceding additive hazard specification. Table 1 shows a hypothetical example in which all the observed dependence between the two events is due to heterogeneity $[r(i, t)]$ that is assumed to enter the hazard functions additively. To keep the example simple I have used discrete time. There are two classes of individuals; those in the same class have the same

[6] A common suggestion for dealing with issues of causality in problems of this type is to use simultaneous-equation methods from economics (see Waite and Stolzenberg, 1976; Marini, 1978). Elsewhere (Winship, 1983a) I have criticized this approach.

TABLE 1
Example with Additive Heterogeneity

Structural Hazard

		Time 1	Time 2
Class 1	Event A	0.51	0.01
	Event B	0.01	0.51
Class 2	Event A	0.99	0.49
	Event B	0.49	0.99

Population Distribution for Combined Classes

		Event B		
		Time 1	Time 2	Time 3
Event A	Time 1	0.245	0.379	0.126
	Time 2	0.001	0.002	0.001
	Time 3	0.004	0.124	0.118

Difference between
Prob($A < B$) − Prob($A > B$) = 0.377
Gamma = 0.7012

Empirical Hazard for Event A

	Time 1	Time 2
B has not occurred	0.75	0.015
B has occurred	—	0.25

Empirical Hazard for Event B

	Time 1	Time 2
A has not occurred	0.25	0.515
A has occurred	—	0.75

hazards (top portion of the table). Note that the hazard functions for class 2 are equal to those for class 1 plus 0.48. Below the hazards is the expected distribution for the combined classes; the two classes are assumed to be of equal size. This distribution is calculated by first generating for the two classes the expected proportions for each event at the three times and then within each class deriving the expected distribution of individuals. Since independence of the two events at the class level is assumed, proportions are obtained by taking the outer product of the two vectors representing the proportions for the two events at each time. Then the distribution functions for the two classes are averaged.

As can be seen by the large gamma value, the two events are highly interrelated. Also shown in Table 1 are the empirical hazards for the population. In the case of event a, at time 2 the hazard for individuals for whom event b has occurred is much higher than for those for whom b has not. Similarly, for event b the hazard is higher for individuals for whom a has occurred than for those for whom it has not. Examples have been constructed for the multiplicative and accelerated time models but are not included here (see Winship 1983b).

This example shows how heterogeneity can lead to apparent interdependence. The next two sections of the chapter examine whether it is possible to differentiate between true interdependence and heterogeneity.

Unrestricted Case. In this section I want to show that if heterogeneity enters the hazard in an unrestricted way the population distribution function $f(t_a, t_b)$ can take any form. The implication is that without putting restrictions on the way that heterogeneity enters the hazard it is impossible to distinguish heterogeneity from true interdependence.

It will be easiest to carry out the proof with distribution rather than hazard functions. No generality is lost in doing this. Start by letting $f_a(i, t)$ and $f_b(i, t)$ be, respectively, the density functions for transitions a and b for individuals of type i. Let $f(t_a, t_b)$ be the joint distribution for the two events in the population. I want to show that any $f(t_a, t_b)$ can be decomposed as follows:

$$f(t_a, t_b) = \int f_a(i, t_a) f_b(i, t_b) K(i)\, di$$

where $K(i)$ is, as above, the density function for i. That is, I want to show that under the assumption that the events are independent at the individual level I can produce any $f(t_a, t_b)$ by properly choosing $f_a(i, t)$, $f_b(i, t)$, and $K(i)$.

Without loss of generality let i, t_a, t_b range continuously from zero to plus infinity. Let $Z(t_a, t_b) = i$ be a 1-to-1 onto function from R^{2+}, the nonnegative quadrant of the real plane, to R^{1+}, the nonnegative segment of the real line. The existence of such a function is guaranteed by the fact that R^{2+} and R^{1+} have the same number of elements (see Boas 1960). Since Z is 1-to-1 onto, it will have a well-defined inverse.

Let $f_a^*(t)$ and $f_b^*(t)$ be densities that have all their mass respectively at t_a^* and t_b^*.[7] Now construct, by appropriate methods, individuals of different types such that for individuals of type $i = Z(t_a, t_b)$ transition a takes place at t_a and transition b takes place at t_b.

Let $f_a(i, t) = f_a^*(r(i, t)) = f_a^*(\alpha_i + \delta_i t)$ and $f_b(i, t) = f_b^*(r(i, t)) = f_b^*(\alpha_i + \delta_i t)$. Choose α_i and δ_i so that if $i = Z(t_a, t_b)$ then $\alpha_i + \delta_i t_a = t_a^*$ and $\alpha_i + \delta_i t_b = t_b^*$. Note that this transformation is of the type found in the accelerated failure-time model.

Now $f_a(i, t) = f_a(Z(t_a, t_b), t)$ and $f_b(i, t) = f_b(Z(t_a, t_b), t)$ are densities that respectively have all their mass at t_a and t_b for $i = Z(t_a, t_b)$. Thus associated with each pair (t_a, t_b) are individuals of type $i = Z(t_a, t_b)$ for whom transitions a and b take place, respectively, at t_a and t_b.

To construct $f(t_a, t_b)$ choose $K(i) = K(Z(t_a, t_b))$ so that there are the correct number of individuals of each type. This is done by letting $K(i) = K(Z(t_a, t_b)) = f(t_a, t_b)$. Then it is the case that for fixed t_a' and t_b':

$$f(t_a', t_b') = \int_0^{+\infty} f_a(i, t_a') f_b(i, t_b') K(i)\, di$$

I have shown that one can construct individual types so that individuals of the same type have the same values for t_a and t_b. Moreover, I have demonstrated that any pair (t_a, t_b) can be associated with a set of individuals for whom transitions a and b take place at t_a and t_b, respectively. Then I have simply chosen $K(i)$ so that there is the correct number of individuals associated with each pair (t_a, t_b) to construct the density function $f(t_a, t_b)$.

Additive Hazards. In this section I want to show that if heterogeneity enters the hazard additively, there are important restrictions on the form of the population joint distribution for the two transitions, $f(t_a, t_b)$. As a result, it is possible to test the null hypothesis that the observed relationship between two transitions arises solely because of heterogeneity in the rates at which individuals mature.

To consider the effects of additive heterogeneity I need to shift from hazard functions and consider survivor functions. For transitions a and b

[7] Technically these two functions are known as Dirac delta functions (Kaplan 1973) and are not proper densities.

the survivor functions are

$$S_a(t_1) = \text{Prob}(t_a > t_1) = \exp\left[-\int_0^{t_1} h_a(t)\,dt\right]$$

$$S_b(t_2) = \text{Prob}(t_b > t_2) = \exp\left[-\int_0^{t_2} h_b(t)\,dt\right]$$

The survivor functions, S_a and S_b, simply equal 1 minus the cumulative density functions, $F_a(t_1)$ and $F_b(t_2)$. The survivor function gives the probability that a transition has not occurred by time t. It is the probability that an individual has "survived" until time t.

Survivor functions have the nice property that they are additive. That is, the population survivor function is simply the integral of the individual survivor functions:

$$S(t) = \int_0^{+\infty} S(i, t) K(i)\,di$$

This property will prove important in developing a test for interdependence. Let $S(i, t_1, t_2)$, the joint survivor function, be the probability that individuals of type i have not experienced transition a before t_1 and transition b before t_2. For a model with heterogeneity but independent transitions, $S(i, t_1, t_2)$ equals the product of the two individual survivor functions:

$$S(i, t_1, t_2) = S_a(i, t_1) S_b(i, t_2)$$

and the population survivor function is

$$S(t_1, t_2) = \int_0^{+\infty} S_a(i, t_1) S_b(i, t_2) K(i)\,di$$

Consider the case where there is additive heterogeneity in the hazards—for example, $h_a(i, t) = u_a(t) + r(i, t)$. Then the survivor functions for transitions a and b are

$$S_a(i, t_1) = \exp\left\{-\int_0^{t_1} [u_a(t) + r(i, t)\,dt]\right\}$$

$$= \left\{\exp\left[-\int_0^{t_1} u_a(t)\,dt\right]\right\}\left\{\exp\left[-\int_0^{t_1} r(i, t)\,dt\right]\right\}$$

$$S_b(i, t_2) = \exp\left\{-\int_0^{t_2} [u_b(t) + r(i, t)]\,dt\right\}$$

$$= \left\{\exp\left[-\int_0^{t_2} u_b(t)\,dt\right]\right\}\left\{\exp\left[-\int_0^{t_2} r(i, t)\,dt\right]\right\}$$

Thus the joint survivor function for individuals of type i can be written

$$S(i, t_1, t_2) = \left\{\exp\left[-\int_0^{t_1} u_a(t)\,dt\right]\right\}$$
$$\times\left(\left\{\exp\left[-\int_0^{t_1} r(i, t)\,dt\right]\right\} \times \left\{\exp\left[-\int_0^{t_2} r(i, t)\,dt\right]\right\}\right)$$
$$\times\left\{\exp\left[-\int_0^{t_2} u_b(t)\,dt\right]\right\}$$

Or letting $R(t_1)$, $G(i, t_1, t_2)$, and $C(t_2)$ denote, respectively, the terms to the left of, within, and to the right of the middle expression:

$$S(i, t_1, t_2) = R(t_1)G(i, t_1, t_2)C(t_2)$$

The only term that depends on i is G:

$$G(i, t_1, t_2) = g(i, t_1)g(i, t_2)$$

where

$$g(i, t_s) = \exp\left[-\int_0^{t_s} r(i, t)\,dt\right]$$

Note that $G(i, t_1, t_2)$ is a symmetric function: $G(i, t_1, t_2) = G(i, t_2, t_1)$. I can now write the population joint survivor function as

$$S(t_1, t_2) = \int_0^{+\infty} R(t_1)G(i, t_1, t_2)C(t_2)K(i)\,di$$

which can be factored into

$$S(t_1, t_2) = R(t_1)\left[\int_0^{+\infty} G(i, t_1, t_2)K(i)\,di\right]C(t_2)$$

Let $G^*(t_1, t_2)$ be the term in the brackets in the last equation so that

$$S(t_1, t_2) = R(t_1)G^*(t_1, t_2)C(t_2)$$

Since $G^*(t_1, t_2)$ is an integral of symmetric functions, $G(i, t_1, t_2)$, $G^*(t_1, t_2)$ is also a symmetric function. Thus the last equation states that $S(t_1, t_2)$ can be written as a product of a function of t_1 alone, $R(t_1)$, a symmetric function in t_1 and t_2, $G^*(t_1, t_2)$, and a function of t_2 alone, $C(t_2)$. Not all joint survivor functions can be written in this form. Borrowing a term from the log-linear model literature (see Bishop, Feinberg, and Holland 1975), I call a survivor function that can be written in this form quasi-symmetric.

This result suggests a test for event interdependence. Specifically, given an estimate of the population joint survivor function one can test the null hypothesis that this function is quasi-symmetric. If the null hypothesis cannot be rejected, then there is evidence that the observed relation between two transitions might solely be the result of heterogeneity.[8] The next section develops such a test.

TESTING FOR QUASI-SYMMETRY

Estimation. Two problems need to be addressed in developing a test for quasi-symmetry. First, in most data sets the survivor function is observed only at a finite number of discrete points. In the data analyzed in the next section, for instance, age of leaving school and age of marriage are measured only in years of age. This approach causes no great problems. It means that the survivor function can be tested for quasi-symmetry at only a discrete set of points. Thus the condition for quasi-symmetry given in the last section becomes

$$\mathbf{S} = \mathbf{R}\mathbf{G}^*\mathbf{C}$$

where, if there are t points of time, $\mathbf{S}$ is a $t \times t$ matrix with the (t_1, t_2) cell being equal to $S(t_1, t_2)$. Further, $\mathbf{R}$ is a $t \times t$ diagonal matrix of row effects with the t_1 diagonal element being equal to $R(t_1)$. Similarly, $\mathbf{C}$ is a $t \times t$ diagonal matrix of column effects. Finally, $\mathbf{G}^*$ is a $t \times t$ symmetric matrix with the t_1, t_2 cell being equal to $G^*(t_1, t_2)$. Thus when a survivor function is observed at a finite number of points one ends up testing whether the realization of that survivor function as a matrix can be decomposed into a set of row, column, and symmetry effects.[9]

This condition looks more familiar described in log-linear notation. The quasi-symmetry condition is

$$\ln \mathbf{S}_{jk} = \mathbf{R}_j^* + \mathbf{C}_k^* + \mathbf{G}_{jk}^{**} \tag{4}$$

[8] Some quasi-symmetric survivor functions cannot be produced from additive heterogeneity. When R, C, and G^* are discrete, a sufficient condition for a quasi-symmetric S to have resulted from additive heterogeneity is that the matrix $\mathbf{G}^*$ be nonnegative definite and that the nonnormalized eigenvectors of $\mathbf{G}^*$ be transformable into a (nonorthogonal) basis such that for each basis vector v the elements of Rv and vC are nonnegative and monotonically decreasing. It is not known whether this is also a necessary condition. I would expect this condition to hold in most applications.

[9] The reader should not think that I have suddenly moved from a continuous to a discrete-time model. The procedure used here assumes that the underlying process occurs in continuous time. The test, however, assumes that this continuous process is observed only at discrete points of time.

with

$$\sum_j \mathbf{G}_{jk}^{**} = \sum_k \mathbf{G}_{jk}^{**} = 0$$

and

$$\mathbf{G}_{jk}^{**} = \mathbf{G}_{kj}^{**} \qquad \text{for all } j, k \tag{5}$$

where $\mathbf{R}_j^* = \ln \mathbf{R}_{jj}$, $\mathbf{C}_k^* = \ln \mathbf{C}_{kk}$, and $\mathbf{G}_{jk}^{**} = \ln G_{jk}^*$.[10] Note that only the last restriction ($\mathbf{G}_{jk}^{**} = \mathbf{G}_{kj}^{**}$) is testable. The other equations are definitions. Thus testing whether a survivor function is quasi-symmetric is equivalent to testing whether $\mathbf{G}_{jk}^{**} = \mathbf{G}_{kj}^{**}$.

The fact that individuals are not identically distributed presents somewhat greater problems. Because of this, I cannot simply estimate restricted and unrestricted forms of the model by maximum likelihood and use a standard likelihood ratio test to test whether $\mathbf{G}_{jk}^{**} = \mathbf{G}_{kj}^{**}$. One approach to the problem is to try to carry out the analysis via marginal-likelihood analysis. I have investigated this approach, and it does not appear possible to derive general results using this method.

An alternative is to carry out the analysis via maximum likelihood or what is more correctly termed quasi-maximum likelihood and to use a Wald or Lagrange multiplier test (Wald 1943; Silvey 1975). These tests are valid even when individuals are not identically distributed (White 1982). I develop a Wald test below.

Start by considering how to get estimates of $\mathbf{R}^*$, $\mathbf{C}^*$, and $\mathbf{G}^{**}$. This can be done by using a maximum-likelihood type of procedure. Consider the matrix $\mathbf{S}$ of survival probabilities. Associated with any matrix $\mathbf{S}$ I can define a matrix $\mathbf{P}$ of the same dimension, where P_{jk} is the proportion of individuals in the population for whom transition a occurs between $j-1$ and j and transition b occurs between $k-1$ and k. The relationship between $\mathbf{P}_{jk}$ and $\mathbf{S}$ is

$$\mathbf{P}_{jk} = \mathbf{S}_{j-1,k-1} - \mathbf{S}_{j-1,k} - \mathbf{S}_{j,k-1} + \mathbf{S}_{j,k} \tag{6}$$

The matrix $\mathbf{S}$ will be a proper survivor function if and only if $\mathbf{P}$ is a proper distribution function—that is, if and only if every cell of $\mathbf{P}$ is nonnegative. In the analysis I impose the further restriction that it be nonzero. This latter restriction is maintained by maximizing the log likelihood. Since the

[10] I have purposely not included a grand mean in Equation (4) in order to keep the notation parallel with that presented above. This means that the row and column effects are identified only up to a constant.

log of zero is minus infinity, estimates of **P** that contain zero values cannot represent maxima.

Think of $\mathbf{P}_{jk}$ as the (pseudo) likelihood for an individual that the two transitions occur in the specified time intervals. It is not the actual likelihood since it does not take into account differences across individuals in their likelihood of having transitions occur at different times. Put another way, this representation does not model the fact that individuals do not have the same probabilities of ending up in each cell of the matrix **P**. Define the pseudo log likelihood of the sample as

$$\mathscr{L} = \sum_i \sum_j \sum_k A_{ijk} \ln \mathbf{P}_{jk} \tag{7}$$

where $A_{ijk} = 1$ if for person i transition a occurs between $j - 1$ and j and transition b occurs between $k - 1$ and k; otherwise $A_{ijk} = 0$. Here the $\mathbf{P}_{jk}$ are functions of the $\mathbf{S}_{jk}$, which are in turn functions of $\mathbf{R}^*$, $\mathbf{G}^{**}$, and $\mathbf{C}^*$. White (1982) shows that maximization of the pseudolikelihood in this situation gives consistent estimates of the population values of **P**. Mosimann (1962) shows that this is true in the specific case of the multinomial. The situation is parallel to that in regression analysis where ordinary least squares gives consistent estimates of the slope parameters even with heteroscedastic errors (Hoadley 1971).

This approach can be used to estimate both restricted ($\mathbf{G}_{jk}^{**} = \mathbf{G}_{kj}^{**}$) and unrestricted forms of the model. In the unrestricted case, estimation is straightforward. The model is just identified and, as a result, the estimated joint survivor function is equivalent to the observed joint survivor function. The parameters can be obtained by solving Equations (4) and ignoring Equations (5).

With the restrictions, the model is overidentified subject to the set of binding constraints defined by Equations (5). A search procedure needs to be used to obtain the parameter values that maximize the likelihood. In the empirical analyses reported later I have used the method of scoring (Rao 1973) to maximize Equation (7) subject to the conditions in Equations (4), (5), and (6) and the constraint that $\mathbf{P}_{jk} > 0$ for all j, k.

Wald Test. One can test whether the data have a quasi-symmetric survivor function by estimating the unrestricted model and then testing whether the symmetry restrictions $\mathbf{G}_{jk}^{**} = \mathbf{G}_{kj}^{**}$ hold. These restrictions can be tested by determining whether the differences $\mathbf{G}_{jk}^{**} - \mathbf{G}_{kj}^{**}$ are significantly different from zero. The procedure is analogous to a t—test in regression analysis where a particular coefficient parameter is estimated without restriction and one then tests whether that estimate is significantly different from zero. The t—test in regression is an example of a Wald test. As with a t—test, the Wald test described here relies on the fact that the parameter estimates are asymptotically distributed multivariate normal (White 1982).

Because individuals are not identically distributed, the usual estimate of the covariance matrix as the inverse of the Fisher information matrix is not consistent. This matrix, however, does provide an upper bound for the covariance matrix of the parameters. (See Domowitz and White 1982, for the general result and Mosimann's 1962 decomposition discussion for the case of the multinomial.) As a result, using this matrix in the Wald test provides a conservative test of the null hypothesis of no interdependence. The formula for the jkth element of the Fisher information matrix is

$$\mathbf{I}_{jk} = \sum_{n} \frac{\partial \log \mathcal{L}}{\partial \theta_j} \frac{\partial \log \mathcal{L}}{\partial \theta_k} \qquad \text{(Fisher information matrix)}$$

where n is the sample size. This expression is also equal to the sample estimate of n times the covariance of the jth and kth elements of the gradient of the log likelihood (the vector of partial derivatives of the log likelihood).

The essential idea behind the Wald test is that if a vector of k variables, v, has a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}$, then the quadratic form $v'\mathbf{\Sigma}^{-1}v$ is distributed as χ^2 with k degrees of freedom (for example, see Hogg and Craig, 1970). If v has a nonzero mean, then this will not be the case. This is the analog to doing a multivariate t—test.

Using this result, the symmetry restriction can be tested by letting $v_k = \mathbf{G}^{**}_{jk} - \mathbf{G}^{**}_{kj}$. Since the $\hat{\mathbf{G}}^{**}$ are distributed asymptotically multivariate normal, the v_k are also distributed asymptotically multivariate normal. The test of the null hypothesis that the data are quasi-symmetric is then carried out by testing the equivalent hypothesis that the vector v has mean zero. As just described, this is done by calculating a χ^2 value from $v'\mathbf{\Sigma}^{-1}v$, where $\hat{\mathbf{\Sigma}}$ is estimated by the information matrix. Very large values relative to the degrees of freedom indicate rejection of the hypothesis that $v = 0$, that is, the hypothesis that the data are quasi-symmetric.

EMPIRICAL ANALYSIS

Data. It is beyond the scope of this chapter to carry out a detailed empirical analysis of the relationship between different life-cycle transitions. I have, however, done a limited analysis of two transitions, the age of leaving school and age of marriage, using the Occupational Change in a Generation II Survey (OCG II). A detailed description of this data set can be found in Featherman and Hauser (1978). Hogan (1978, 1981) has carried out detailed analyses of these two variables with the same data.

The sample analyzed is restricted in several ways. First, OCG II is limited to males. Second, although a survivor function can be estimated with right censoring at any age (Kaplan and Meier, 1958), I have restricted the sample to individuals aged 30 or older and have analyzed the joint survivor function only up to age 30. By this age most males have left school and married, so this is not a particularly burdensome restriction. With the hope of minimizing the heterogeneity in the data, I have also limited the analysis to nonblacks.

The next section provides a descriptive analysis of the data. The point is to show that at a simple level there does seem to be some interdependence between the age of leaving school and age of marriage. The following section describes the results of carrying out the Wald test for quasi-symmetry. Then the quasi-symmetry model is fit to the data and the fitted model is compared to the actual data. The analysis is then repeated for individuals who leave school and marry after the age of 19.

Descriptive Analysis. Table 2 shows the frequency distribution for the two events for a total of 17,230 individuals. Only 288 individuals have neither married nor finished school by age 30. Gamma for this table is 0.11722, indicating a modest association between the two events. Individuals are much more likely to leave school before marrying than the reverse; the difference between the proportions of the population following each of these patterns is 0.5618. These findings are consistent with those of Hogan (1978, 1981).

Tables 3 and 4 show, respectively, the hazard rates for marriage and schooling broken out in several ways. At the top of Table 3 is the simple marital hazard rate by age. Here the hazard is simply the proportion of the population who have not yet married who marry in a given year. As one might expect, the hazard increases with age and then levels out toward 30. The second and third rows of Table 3 show the marital hazard by school enrollment status. At all ages the marital hazard is

TABLE 2

Observed Distribution of Nonblacks Over 30 Years of Age

Age of Leaving School	Age of Marriage																	
	≤ 13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	≥ 30	Total
≤ 13	23	5	14	15	42	62	78	93	96	90	76	72	61	63	55	42	36	166	1,089
14	3	1	3	12	13	35	56	58	72	71	60	46	48	50	32	29	17	110	716
15	4	0	3	12	26	44	81	110	86	79	72	58	62	47	46	30	25	123	908
16	1	1	4	20	48	72	127	159	143	150	113	92	102	78	61	40	39	160	1,410
17	2	0	1	7	48	87	127	161	209	176	169	112	111	87	66	62	42	155	1,622
18	6	0	2	5	56	154	243	305	347	342	267	247	186	155	108	107	63	280	2,873
19	2	1	3	3	7	73	182	213	226	255	214	230	163	113	83	72	52	198	2,090
20	2	0	0	1	7	13	52	118	136	131	116	91	81	64	69	40	38	98	1,057
21	0	0	1	0	0	11	11	52	96	85	67	64	64	44	26	31	20	91	663
22	1	1	0	1	4	5	22	31	78	91	66	56	73	43	35	35	17	77	636
23	0	1	0	1	6	13	15	29	48	59	83	61	49	47	25	25	23	67	552
24	0	1	0	1	1	3	16	19	34	54	59	63	42	43	27	22	20	57	462
25	0	0	0	3	2	7	12	22	47	50	39	41	56	52	26	23	14	50	444
26	0	0	0	1	0	4	11	30	35	42	36	36	38	36	29	24	15	47	384
27	1	0	0	1	1	6	12	21	18	44	32	25	26	25	22	17	19	42	312
28	0	0	0	0	2	7	9	17	23	43	26	25	29	14	14	17	9	37	272
29	1	0	0	1	2	4	8	13	24	24	29	27	13	21	16	10	17	30	240
≤ 30	3	1	2	2	8	30	71	78	117	163	151	139	127	94	80	82	64	288	1,500
Total	49	12	33	86	273	630	1,133	1,529	1,835	1,949	1,675	1,485	1,331	1,076	820	708	530	2,076	17,230

Note: Gamma = 0.11722; Prob($A < B$) − Prob($B < A$) = 0.5618.

TABLE 3
Marital Hazard

		Respondent's Age															
		14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Total Population																	
		0.001	0.002	0.005	0.016	0.038	0.070	0.102	0.136	0.167	0.173	0.185	0.203	0.207	0.198	0.214	0.203
By School Enrollment Status																	
In School		0.000	0.001	0.003	0.011	0.029	0.050	0.070	0.108	0.149	0.160	0.170	0.190	0.178	0.173	0.197	0.203
Out of School		0.005	0.010	0.015	0.032	0.055	0.091	0.124	0.152	0.176	0.178	0.190	0.208	0.214	0.204	0.217	0.203
By Age of Leaving School																	
Age of Leaving School	≤ 13	0.005	0.013	0.014	0.041	0.063	0.084	0.109	0.127	0.136	0.133	0.145	0.144	0.174	0.184	0.172	0.178
	14	**0.001**	0.004	0.017	0.019	0.051	0.086	0.098	0.135	0.153	0.153	0.139	0.168	0.210	0.170	0.186	0.134
	15	0.000	**0.003**	0.013	0.029	0.051	0.099	0.149	0.137	0.146	0.156	0.148	0.186	0.173	0.205	0.169	0.169
	16	0.001	0.003	**0.014**	0.035	0.054	0.100	0.140	0.146	0.180	0.165	0.161	0.212	0.206	0.203	0.167	0.196
	17	0.000	0.001	0.004	**0.030**	0.056	0.086	0.119	0.176	0.180	0.210	0.176	0.212	0.211	0.203	0.239	0.213
	18	0.000	0.001	0.002	0.020	**0.055**	0.092	0.127	0.165	0.195	0.189	0.216	0.207	0.217	0.194	0.238	0.184
	19	0.000	0.001	0.001	0.003	0.035	**0.091**	0.117	0.141	0.185	0.190	0.252	0.239	0.218	0.205	0.224	0.208
	20	0.000	0.000	0.001	0.007	0.012	0.050	**0.120**	0.157	0.180	0.194	0.189	0.208	0.207	0.282	0.227	0.279
	21	0.000	0.002	0.000	0.000	0.017	0.017	0.081	**0.163**	0.173	0.165	0.188	0.232	0.208	0.155	0.218	0.180
	22	0.002	0.000	0.002	0.006	0.008	0.035	0.051	0.137	**0.185**	0.164	0.167	0.261	0.208	0.213	0.271	0.181
	23	0.002	0.000	0.002	0.011	0.024	0.028	0.056	0.099	0.134	**0.218**	0.205	0.208	0.251	0.179	0.217	0.256
	24	0.002	0.000	0.002	0.002	0.007	0.035	0.043	0.081	0.140	0.177	**0.230**	0.199	0.254	0.214	0.222	0.260
	25	0.000	0.000	0.007	0.005	0.016	0.028	0.052	0.118	0.142	0.130	0.156	**0.253**	0.315	0.230	0.264	0.219
	26	0.000	0.000	0.003	0.000	0.010	0.029	0.082	0.104	0.139	0.138	0.160	0.201	**0.238**	0.252	0.279	0.242
	27	0.000	0.000	0.003	0.003	0.019	0.040	0.072	0.067	0.175	0.154	0.142	0.172	0.200	**0.220**	0.218	0.311
	28	0.000	0.000	0.000	0.007	0.026	0.034	0.067	0.097	0.201	0.152	0.172	0.242	0.154	0.182	**0.270**	0.196
	29	0.000	0.000	0.004	0.008	0.017	0.034	0.058	0.114	0.128	0.178	0.201	0.121	0.223	0.219	0.175	**0.362**
	≥ 30	0.001	0.001	0.001	0.005	0.020	0.049	0.056	0.090	0.137	0.147	0.159	0.173	0.155	0.156	0.189	0.182

TABLE 4
Schooling Hazard

		Respondent's Age															
		14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Total Population																	
		0.044	0.059	0.097	0.124	0.250	0.243	0.162	0.121	0.132	0.133	0.128	0.141	0.142	0.134	0.135	0.138
By Marital Status																	
Not Married		0.044	0.059	0.097	0.124	0.247	0.239	0.160	0.122	0.129	0.134	0.131	0.145	0.141	0.131	0.114	0.118
Married		0.115	0.138	0.158	0.169	0.476	0.353	0.187	0.117	0.145	0.130	0.123	0.137	0.142	0.136	0.143	0.144
By Age of First Marriage																	
Age of First Marriage	≤ 13	0.115	0.174	0.053	0.111	0.375	0.200	0.250	0.000	0.167	0.000	0.000	0.000	0.000	0.200	0.000	0.250
	14	**0.143**	0.000	0.167	0.000	0.000	0.200	0.000	0.000	0.250	0.333	0.500	0.000	0.000	0.000	0.000	0.000
	15	0.158	**0.188**	0.308	0.111	0.250	0.500	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	16	0.169	0.203	**0.426**	0.259	0.250	0.200	0.083	0.000	0.091	0.100	0.111	0.375	0.200	0.250	0.000	0.333
	17	0.056	0.119	0.250	**0.333**	0.583	0.175	0.212	0.000	0.154	0.273	0.063	0.133	0.000	0.077	0.167	0.200
	18	0.062	0.083	0.147	0.209	**0.467**	0.415	0.126	0.122	0.063	0.176	0.049	0.121	0.078	0.128	0.171	0.118
	19	0.053	0.081	0.138	0.161	0.366	**0.432**	0.218	0.059	0.125	0.097	0.115	0.098	0.099	0.120	0.102	0.101
	20	0.040	0.080	0.125	0.145	0.322	0.331	**0.274**	0.167	0.119	0.127	0.095	0.122	0.189	0.163	0.157	0.143
	21	0.041	0.052	0.090	0.145	0.282	0.256	0.207	**0.185**	0.184	0.139	0.114	0.178	0.161	0.099	0.140	0.170
	22	0.038	0.044	0.088	0.113	0.247	0.245	0.167	0.130	**0.160**	0.123	0.129	0.137	0.133	0.161	0.187	0.128
	23	0.038	0.047	0.077	0.125	0.225	0.233	0.165	0.114	0.127	**0.182**	0.159	0.125	0.131	0.134	0.126	0.161
	24	0.033	0.042	0.070	0.092	0.224	0.268	0.145	0.119	0.118	0.146	**0.177**	0.140	0.143	0.116	0.131	0.163
	25	0.038	0.051	0.088	0.105	0.196	0.214	0.135	0.124	0.161	0.129	0.127	**0.194**	0.163	0.133	0.172	0.093
	26	0.049	0.049	0.085	0.104	0.206	0.190	0.133	0.105	0.115	0.142	0.151	0.215	**0.189**	0.162	0.109	0.183
	27	0.042	0.063	0.089	0.105	0.193	0.184	0.187	0.087	0.128	0.105	0.126	0.139	0.180	**0.167**	0.127	0.167
	28	0.044	0.047	0.066	0.109	0.212	0.181	0.123	0.108	0.137	0.114	0.113	0.133	0.160	0.135	**0.156**	0.109
	29	0.034	0.052	0.086	0.102	0.170	0.169	0.148	0.092	0.086	0.127	0.127	0.101	0.121	0.174	0.100	**0.210**
	≥ 30	0.058	0.068	0.095	0.102	0.206	0.183	0.111	0.116	0.111	0.108	0.103	0.101	0.106	0.106	0.104	0.094

higher for individuals not in school than for those who are. The difference between the hazards for the two groups, however, is much larger at younger ages and declines with age so that by the late twenties there is almost no difference at all.

The rest of Table 3 shows the marital hazard by age of leaving school. The hazard rate for the year in which an individual leaves school is shown in boldface type. If one believes that when an individual leaves school his chances of marrying should increase, then the rates to the right of the emphasized rate should be higher than those to the left. Although the pattern is far from being clean, especially at the youngest ages of leaving school, this in general is the case. An example is individuals who leave school at age 23, for whom the marital hazard at age 22 is 0.134 and at age 24 is 0.205.

Table 4 shows the same set of results for age of leaving school. At the top of the table is the hazard for the whole sample. In this case the hazard rises quickly to peak at ages 18 and 19 when individuals are finishing high school and then declines. The next two rows show the hazard by marital status. At the youngest ages married individuals are much more likely to leave school than individuals who are not married. After age 21, though, marital status appears to make no difference.

The bottom part of Table 4 shows the hazard for leaving school by age of first marriage. Analogous to Table 3, the hazards for the year in which an individual first marries have been emphasized. As in Table 3, one might expect that individuals would be more likely to leave school after they have married. At the youngest ages this seems to be the case. For instance, the hazard for individuals who marry at age 17 is 0.250 at age 16 but 0.583 at age 18. For individuals who marry at age 19 or later, the pattern is mixed, though in most cases it is the opposite of that at youngest ages. That is, individuals are less likely to leave school after they marry than before. For example, individuals who marry at age 26 have at age 25 a hazard of 0.215 and at age 27 a hazard of 0.162.

These simple analyses suggest that there is interdependence between the age of leaving school and age of marriage. The timing of these two events is modestly correlated and individuals are very likely to leave school before they marry. Moreover, an examination of the hazards suggests interdependence. In particular, leaving school appears to increase the marital hazard. This effect is greatest at the youngest ages.

Wald Test. The Wald test for quasi-symmetry indicates whether the observed association between age of leaving school and age of marriage could be explained simply by additive heterogeneity. The test is carried out assuming that individuals differ only in terms of a single additive component in their hazards. Thus the test does not allow the relationship between the age of leaving school and age of marriage to differ across individuals. More detailed analyses would let this relation vary as a function of a set of observed variables, X. Such an analysis must await the future.

To carry out the Wald test for quasi-symmetry I had to collapse the data. In order to eliminate zero cells, I have collapsed the categories for ages less than or equal to 13 to 17 into a single category. This procedure also gives a matrix of manageable size. Even so, the Wald test involves inverting a 169×169 matrix, close to the limit of the computer system used (a Cyber 170-730). The collapsed data behave very much like the full table. Gamma is 0.1324, somewhat higher than in the original table. A standard χ^2 test for independence using the log-likelihood form of the χ^2 statistic (Bishop, Feinberg, and Holland 1975) gives a χ^2 of 1,175.82 with 169 degrees of freedom, indicating a highly significant lack of independence.[11] In the collapsed table individuals are still much more likely to leave school before marrying than the reverse; the difference between the two proportions is 0.55212.[12]

The row, column, and cell parameters needed in the Wald test were found by solving Equations (4). This can be done simply by forming the log of the empirical joint survivor function and then using an ANOVA type procedure whereby the row and column effects are taken out leaving the cell effects. The Wald test was calculated by a program written in APL. The result was a χ^2 statistic of 378.30 with 78 degrees of freedom. This result is highly significant and represents a resounding rejection of the quasi-symmetry hypothesis. Using the normal approximation to the χ^2, this is a little over 15 standard deviations from the mean.

Although the quasi-symmetry model as a whole fits very poorly, it is important to see whether it fits equally poorly over the whole age range. To answer this question I looked at the ratio of the differences $G_{jk}^{**} - G_{kj}^{*}$

[11] All statistical tests reported in this chapter are based on the assumption that OCG II is a simple random sample. In fact, it is a cluster sample. What the relative size of the design effects are for the models tested is unclear. It is quite doubtful that such corrections would affect the quality of the results reported.

[12] I conducted detailed analyses of the survivor function at younger ages. These analyses confirm the result reported in the chapter that at younger ages the quasi-symmetry model fits the data quite poorly.

to their standard errors. Looking at these results (not presented here), a definite pattern emerges. First, terms involving ages 17 or less, 18, and 19 are those most often likely to be significantly different from zero. Second, relative to the number of individuals who left school at age 17 or less, 18, or 19 and then married shortly thereafter, there are too few individuals in the data relative to the model estimates who married at these ages and then left school thereafter. Putting it the other way around, at the youngest ages there are too many individuals who leave school and then marry soon after. This finding is consistent with the hazard rate analysis reported above.

To examine this matter further, I fit the quasi-symmetry model to the collapsed data in order to compare the observed data with those predicted by the quasi-symmetry model. To do this it was necessary to collapse the data further by grouping categories for ages 18 through 29 into adjacent pairs giving an 8×8 matrix. A maximum-likelihood routine based on the method of scoring (Rao, 1973) written in APL was used. The process of estimation itself suggested how poorly the model fit the data. It was never possible to get the program to converge. After 634 iterations the search procedure had not stopped, though the parameter values were changing less than 0.000004 per iteration. I therefore decided to use the parameter values at the 634 iteration as the estimates.

Table 5 shows the ratio between the observed and expected frequencies. Several results should be noted. In the first column there are too few observed relative to expected individuals in all cells. This can happen

TABLE 5
Observed Divided by Expected Frequencies: Quasi-Symmetry Model

Age of Leaving School	Age of Marriage							
	≤ 17	18–19	20–21	22–23	24–25	26–27	28–29	≥ 30
≤ 17	0.93	1.17	1.17	0.93	0.87	0.98	0.89	1.08
18–19	0.53	0.90	1.18	1.11	1.09	0.93	0.89	0.88
20–21	0.32	0.48	0.91	1.07	1.10	1.08	1.05	1.01
22–23	0.76	0.64	0.83	1.01	1.10	1.07	1.03	1.01
24–25	0.90	0.66	0.81	0.93	1.04	1.21	1.07	0.98
26–27	0.53	1.24	0.85	0.96	0.87	1.05	1.19	1.08
28–29	0.71	2.08	0.93	1.03	1.00	0.92	1.07	1.02
≥ 30	0.54	1.41	0.92	1.01	1.04	0.98	1.07	1.02

because the model constrains the row and column marginals of the survivor function, not the distribution.[13] Second, the first row of the table (minus the first element) seems to suggest a pattern. There are too many observed relative to expected individuals who marry right after leaving school as opposed to marrying later. Again, this pattern suggests that there is some interdependence. If one accepts Hogan's (1978, 1981) proposition that norms govern the sequencing of transitions, one might argue that the data support a hypothesis that there is a norm of not marrying before finishing high school. An examination of the second row from the third column on reveals this same type of pattern.[14]

Since the analysis suggests that the lack of fit of these data occurs mostly at the younger ages, I decided to test the survivor function for quasi-symmetry between the ages of 20 and 30. This test is equivalent to examining individuals who left school and married after the age of 19. Doing so reduces the sample size to 6,120. Gamma for this subtable is 0.1010 and the standard χ^2 test for independence gives a result of 251.33 with 100 degrees of freedom. The difference between the proportion of individuals who left school before marrying from the proportion who did the reverse is only 0.0273. A Wald statistic was calculated to test the quasi-symmetry hypothesis for this portion of the survivor function. This calculation results in a χ^2 of 49.1 with 45 degrees of freedom. Under the assumption of a simple random sample, this χ^2 has a probability of 0.7151, indicating that quasi-symmetry cannot be rejected for this component of the survivor function.

The quasi-symmetry model for this subtable was also estimated. Convergence was achieved in 20 iterations. Examination of the ratio of observed to expected frequencies (not presented) indicates that the model fits the data quite well. If anything is to be made of the pattern of fit, it is just the reverse of that for younger age groups presented earlier: There are too few individuals marrying right after finishing school. Hence both the Wald statistic and the size of the discrepancy between the observed and

[13] One reviewer suggested constraining the row and column marginals of the distribution rather than those of the survivor function. As discussed above, the quasi-symmetry model implies that it is the row and column marginals of the survivor function that are constrained.

[14] Allowing parameters of the quasi-symmetry model to vary with respect to observed X's might well increase the fit of the model to the data. This approach needs to be investigated in future analyses. The hypothesis that there is a structural relationship between age of leaving school and age of marriage such that individuals wait to marry until after finishing high school, however, seems quite plausible.

expected frequencies suggest that the quasi-symmetry model fits this portion of the data well.[15]

In summary, then, the descriptive analyses and tests of the quasi-symmetry hypothesis have provided several results. At the youngest ages both the descriptive analysis and the Wald test suggest that there may be interdependence between the age of leaving school and age of marriage. In particular, leaving school appears to increase the hazard for marriage. At older ages the descriptive analysis suggests that there is interdependence, but it is much weaker. A test of the quasi-symmetry hypothesis, however, suggests that all of this interdependence may be explained by unobserved heterogeneity.

CONCLUSION

This chapter has examined two possible reasons why the timing of different life-cycle transitions may be interrelated. There may be a causal relationship or, alternatively, the observed relation may be the result of heterogeneity across individuals in the rates at which they approach adulthood. In colloquial language: Some people bloom late and others bloom early.

I have looked at the question of whether these two possibilities are observationally distinct. In the general case I have shown that they are not. When heterogeneity is only allowed to enter the hazard additively, however, a distinction can be made. A test of the null hypothesis that the relation between two events is due simply to additive heterogeneity was then developed.

Finally, the relation between the age of leaving school and age of marriage was examined using the OCG II data. Fairly simple analyses pointed to interdependence between these two transitions. A test of the null hypothesis that this observed interdependence was due to additive heterogeneity suggested, however, that there was evidence of interdependence only at the youngest ages.

[15] As pointed out in note 8 quasi-symmetry does not guarantee that a survivor function can be obtained from additive heterogeneity. Note 8 indicates an additional sufficient condition. I calculated the eigenvalues and eigenvectors for my estimate of G^*. All but two eigenvalues were within machine error of zero. The first eigenvalue was quite large and represented over 95 percent of the trace; the other eigenvalue represented something less 5 percent of the trace. It was then quite easy to find an oblique rotation of the two eigenvectors associated with these two eigenvalues that gave basis vectors satisfying the condition in note 8.

APPENDIX: OBSERVED X'S

In the main body of the chapter I have made the strong assumption that the relation between transitions a and b is the same across individuals. Mathematically this is equivalent to the functions $R(t_1)$ and $C(t_2)$ not varying across individuals. As has already been noted, this is a strong assumption. Failure of the quasi-symmetry model to fit the data may result from this condition not holding as opposed to true interdependence being present.

One solution to this problem is to let the hazards for the two events be a function of a set of observed variables X. Thus the relation between the two transitions can vary with X. Then the survivor functions for individuals with observed variables X and of type i are

$$S_a[i, t_1, X(\theta)] = \exp\left(-\int_0^{t_1}\{u_a[t, X(\theta)] + r(i,t)\}\, dt\right)$$

$$S_b[i, t_2, X(\phi)] = \exp\left(-\int_0^{t_2}\{u_b[t_2, X(\phi)] + r(i,t)\}\, dt\right)$$

where θ and ϕ are parameters to be estimated. Using a factoring similar to that used earlier in the text, the survivor function for individuals with covariates X is equal to

$$\begin{aligned} S(t_1, t_2, X) = \exp&\left\{ -\int_0^{t_1} u_a[t, X(\theta)]\, dt\right\} \\ &\times \int_0^{+\infty} \exp\left\{ -\int_0^{t_1} r(i,t)\, dt\right\} \\ &\times \exp\left\{ -\int_0^{t_2} r(i,t)\, dt\right\} K(i|X)\, di \\ &\times \exp\left\{ -\int_0^{t_2} u_b[t, X(\phi)]\, dt\right\} \end{aligned}$$

The population survivor function is then just the integral of this expression over all X. Since a sum of quasi-symmetric functions is in general not itself quasi-symmetric, the population survivor function is not in general quasi-symmetric. Notice that, as before, the first term is a function of t_1, the second term of t_1 and t_2, and the third term of t_2. Each term is also a function of X. Using previous notation and leaving θ and ϕ implicit, the preceding expression can be written as

$$S(t_1, t_2, X) = R(t_1, X)G(t_1, t_2, X)C(t_2, X)$$

where, as before, R, G, C are the first, second, and third terms in the foregoing equation. The dependence of G on X is due to the fact that X and i may not be independent. If they are independent, $K(i|X) = K(i)$ and G will not be a function of X.

As in the original formulation, this model can be estimated by pseudo maximum likelihood with the parameters in R, G, C being functions of X. With a large sample and few values for X, the easiest way to do this is to stratify on X and estimate R, G, C for each set of individuals separately. This may be impossible if there are time-varying X's, however, because distinct time paths represent distinct values of X. Since the stratified samples are independent, a Wald test can be carried out on each set of individuals and the χ^2 statistics from each test can be added to produce a χ^2 statistic for the entire population.

With samples that are too small to stratify, the effects of X on R, G, and C have to be parameterized. In this case parameter estimates can be obtained by maximizing the pseudolikelihood for the sample. As before, when we have observations at discrete points in time we can define the pseudolikelihoods for individuals from their survival probabilities. Equation (6) in the main body of the chapter does this. The maximum-likelihood estimates can then be obtained by using a nonlinear optimization program. (This will probably be very expensive.) The Wald test then consists of testing whether the appropriate differences between the parameters of $G(t_1, t_2, X)$ are equal to zero.

Variation in R and C with X can be interpreted as representing differences across groups in the timing of the two transitions and in relative position of the transitions to each other. Variation in G with X is the result of dependence of i on X. Generally, this variation will not be interpretable.

REFERENCES

BISHOP, YVONNE M. M., STEPHEN E. FEINBERG, AND PAUL HOLLAND. 1975. *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.

BOAS, RALPH P. 1960. *A Primer of Real Functions*. New York: Wiley.

COLEMAN, JAMES S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.

DOMOWITZ, IAN AND HALBERT WHITE. 1982. "Misspecified Models with Dependent Observations." *Journal of Econometrics* 20:35–58.

ELDER, GLEN H. JR. 1975. "Age Differentiation and the Life Course." Pp. 165–90 in *Annual Review of Sociology*. Vol. 1, edited by A. Inkeles, J. Coleman, and N. Smelser. Palo Alto, CA: Annual Reviews.

_______________. 1978. "Family History and the Life Course." Pp. 17–64 in *Transitions: The Family and the Life Course in Historical Perspective*, edited by T. K. Hareven. New York: Academic Press.

FEATHERMAN, DAVID L. AND ROBERT HAUSER. 1978. *Opportunity and Change*. New York: Academic Press.

HECKMAN, JAMES J. 1978. "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Departure." *Annals de Insee* 30–31:227–70.

_______________. 1981. "Heterogeneity and State Dependence." Pp. 91–136 in *Conference on Labor Markets*, edited by S. Rosen. Chicago: University of Chicago Press.

HECKMAN, JAMES J. AND GEORGE BORJAS. 1980. "Does Unemployment Cause Future Unemployment?: Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica* 47:247–83.

HECKMAN, JAMES J. AND BURTON SINGER. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52:271–320.

HOADLEY, BRUCE. 1971. "Asymptotic Properties of Maximum Likelihood Estimates for the Independent Not Identically Distributed Case." *Annals of Mathematical Statistics* 42:1977–91.

HOGAN, DENNIS P. 1978. "The Variable Order of Events in the Life Course." *American Sociological Review* 43:573–86.

_______________. 1981. *Transitions and Social Change: The Early Lives of American Men*. New York: Academic Press.

HOGG, ROBERT V. AND ALLEN T. CRAIG. 1970. *Introduction to Mathematical Statistics*. New York: Macmillan.

KALBFLEISH, JOHN D. AND ROSS L. PRENTICE. 1980. *The Statistical Analysis of Failure Time Data*. New York: Wiley.

KAPLAN, EDWARD L. AND PAUL MEIER. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association* 53:457–81.

KAPLAN, WILFRED. 1978. *Advanced Calculus*. Reading, MA: Addison-Wesley.

MARE, ROBERT D., CHRISTOPHER WINSHIP, AND WARREN N. KUBITSCHEK. 1984. "The Transition from Youth to Adult: Understanding the Age Pattern of Employment." *American Journal of Sociology* 90:326–58.

MARINI, MARGARET M. 1978. "The Transitions to Adulthood: Sex Differences in Educational Attainment and Age at Marriage." *American Sociological Review* 43:483–507.

MODELL, JOHN AND TAMARA K. HAREVEN. 1978. "Transitions: Patterns of Timing." Pp. 245–69 in *Transitions: The Family and the Life Course in Historical Perspective*, edited by T. K. Hareven. New York: Academic Press.

MOSIMANN, JAMES E. 1962. "On the Compound Multinomial Distribution. The β-distribution, and Correlations Among Proportions." *Biometrika* 49:65–82.

NEUGARTEN, BERNICE L., JOAN W. MOORE, AND JOHN C. LOWE. 1965. "Age Norms, Age Constraints, and Adult Socialization." *American Journal of Sociology* 70:710–17.

RAO, C. RADHAKRISHNA. 1973. *Linear Statistical Inference and Its Applications* (2nd ed.). New York: Wiley.

RILEY, MATILDA W., MARILYN E. JOHNSON, AND ANNE FONER (EDS.). 1972. *Aging and Society*. Vol. 3, *A Sociology of Age Stratification*. New York: Russell Sage.

SILVEY, SAMUEL D. 1975. *Statistical Inference*. London: Chapman & Hall.

TUMA, NANCY B. 1980. "When Can Interdependence in a Dynamic System of Qualitative Variables Be Ignored?" Pp. 358–91 in *Sociological Methodology 1980*, edited by K. F. Schuessler. San Francisco: Jossey-Bass.

TUMA, NANCY B. AND MICHAEL HANNAN. 1984. *Social Dynamics: Models and Methods*. New York: Academic Press.

VARIAN, HAL R. 1984. *Microeconomic Analysis*. New York: Norton.

WAITE, LINDA J. AND ROSS M. STOLZENBERG. 1976. "Intended Childbearing and Labor Force Participation of Young Women: Insights from Non-recursive Models." *American Sociological Review* 41:235–52.

WALD, ABRAHAM. 1943. "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations in Large." *Transactions of the American Mathematical Society* 54:426–82.

WHITE, HALBERT. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50:1–25.

WINSHIP, CHRISTOPHER. 1983a. "Marriage and Employment: Modelling the Interdependence." Unpublished manuscript. Northwestern University.

_______________. 1983b. "Age Dependence, Heterogeneity, and the Interdependence of Life Cycle Transitions." Paper presented at the annual meetings of the American Sociological Association, Detroit.

WOHLWILL, JOACHIM F. 1970. "The Age Variable in Psychological Research." *Psychological Review* 77:49–64.

YAMAGUCHI, KAZUO. 1983. "A New Approach to the Problem of Unobserved Population Heterogeneity for the Analysis of Repeatable Events in Life Histories." Paper presented at the annual meetings of the American Sociological Association, Detroit.