Duration Models: Specification, Identification, and Multiple Durations

Van den Berg, Gerard J.

VU University Amsterdam

2000

Duration Models: Specification, Identification, and Multiple Durations

Gerard J. van den Berg *

March 6, 2000

*Department of Economics, Free University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, CEPR, and Tinbergen Institute.
E-mail: gberg@econ.vu.nl.

Forthcoming in: James J. Heckman and Edward Leamer, editors, *Handbook of Econometrics, Volume V* (North-Holland, Amsterdam), to appear in 2000.

Thanks to Tony Lancaster, Geert Ridder, Jean-Marc Robin, Jim Heckman, Zvi Eckstein, and Jaap Abbring, for their very useful suggestions. Also thanks to colleagues at the Free University Amsterdam for stimulating discussions.

Abstract

Since the early 1980s, the econometric analysis of duration variables has become widespread. This chapter provides an overview of duration analysis, with an emphasis on the specification and identification of duration models, and with special attention to models for multiple durations. Most of the chapter deals with so-called reduced-form duration models, notably the popular Mixed Proportional Hazard (MPH) model and its multivariate extensions. The MPH model is often used to describe the relation between the empirical exit rate and "background variables" in a concise way. However, since the applications usually interpret the results in terms of some economic-theoretical model, we examine to what extent the deep structural parameters of some important theoretical models can be related to reduced-form parameters. We subsequently examine the specification and identification of the MPH model in great detail, we provide intuition on what drives identification, and we infer to what extent biases may occur because of misspecifications. This examination is carried out separately for the case of single-spell data and the case of multi-spell data. We also compare different functional forms for the unobserved heterogeneity distribution.

Next, we examine models for multiple durations. In the applied econometric literature on the estimation of multiple-duration models, the range of different models is actually not very large. Typically, the models allow for dependence between the duration variables by way of their unobserved determinants, with each single duration following its own MPH model. In addition to this, the model may allow for an interesting "causal" effect of one duration on the other, as motivated by an underlying economic theory. For all these models we examine the conditions for identification. Some of these are intimately linked to particular estimation strategies. The multiple-duration model where the marginal duration distributions each satisfy an MPH specification, and the durations can only be dependent by way of their unobserved determinants, is called the Multivariate Mixed Proportional Hazard (MMPH) model. For this model, we address the issue of the dimensionality of the heterogeneity distribution and we compare the flexibility of different parametric heterogeneity distributions.

On a number of occasions, we incorporate recent insights from the biostatistical literature on duration analysis, and we contrast points of view in this literature to those in the econometric literature. Finally, throughout the chapter, we discuss the importance of the possible collection of additional data.

Contents

continued on next page

1 Introduction

Duration analysis is a core subject of econometrics. Since the early 1980s, the empirical analysis of duration variables has become widespread. There are a number of distinct reasons for this development. First of all, many types of behavior over time tend increasingly to be regarded as movements at random intervals from one state to another. Examples include movements by individuals between the labor market states of employment, unemployment and nonparticipation, and movements between different types of marital status. This development reflects the fact that dynamic aspects of economic behavior have become more important in economic theories, and that in these theories the arrival of new information (and thus the change in behavior in response to this) occurs at random intervals. Secondly, longitudinal data covering more than just one spell per respondent are widely available in labor economics, as well as in demography and medical science. Applications of duration analysis include, in labor economics, the duration of unemployment and the duration of jobs (see e.g. the survey by Devine and Kiefer, 1991), strike durations (e.g., Kennan, 1985), and the duration of training programs (Bonnal, Fougère, and Sérandon, 1997). In business economics, duration models have been used to study the duration until a major investment (e.g., Anti Nilsen and Schiantarelli, 1998). In population economics, duration analysis has been applied to study marriage durations (Lillard, 1993), the duration until the birth of a child (Heckman and Walker, 1990), and the duration until death. In econometric analyses dealing with selective observation, duration models have been used to study the duration of panel survey participation (e.g., Van den Berg and Lindeboom, 1998). In marketing, duration models have been used to study household purchase timing (e.g., Vilcassim and Jain, 1991), in consumer economics to study the duration until purchase of a durable or storable product (Antonides, 1988, and Boizot, Robin and Visser, 1997), and in migration economics to study the duration until return migration (e.g., Lindstrom, 1996). Recently, duration models have been applied in areas in economics where the unit under consideration is not an individual or firm. For example duration models have been used in macro economics to study the duration of business cycles (e.g., Diebold and Rudebusch, 1990), in finance to study the duration between stock-market share transactions (Engle and Russell, 1998), in political economics to study the duration of wars (see Horvath, 1968), and in industrial organization to study the duration of a patent (Pakes and Schankerman, 1984).

This chapter presents an overview of duration analysis. A substantial part of the chapter deals with so-called reduced-form duration models, notably the famous Mixed Proportional Hazard (MPH) model. This model expresses the exit rate to a destination state as a rather simple function of observed and unobserved explanatory variables and the elapsed duration in the current state. This model and its special cases, most notably the Proportional Hazard (PH) model, have been used in hundreds of empirical studies (see e.g. Devine and Kiefer, 1991, for references in micro labor economics). Parametric versions of the model are included in statistical packages like STATA, SAS, S-PLUS and SPSS (see Pelz and Klein, 1996, for a comparison of some packages). We examine the specification and identification of the MPH model in detail, and we infer to what extent biases may occur because of misspecifications.

The MPH model is often used to describe the relation between the empirical exit rate and "background variables" in a concise way, and to provide estimates of the effect of an explanatory variable on the duration variable. However, since the applications usually interpret the results in terms of some economic-theoretical model, it is important to examine to what extent the deep structural parameters of this theoretical model can be related to the reduced-form parameters. As we shall see, economic theory in general does not lead to a "proportional" specification as in the MPH duration model, and this complicates the interpretation of the reduced-form estimates.

Recently, the empirical analysis of multiple durations has become widespread. In many cases it is simply a necessity to address the issue of whether different durations (given the observed explanatory variables) are not independently distributed. For example, if the duration data are censored then it matters for empirical inference how the time until censoring is related to the duration of interest. More generally, if a spell under observation can terminate in a number of different ways ("competing risks") then it matters whether the latent durations to the different destinations are related. As we shall see, economic theory often predicts that such durations are related. In fact, the issue of whether different durations are related is often an important question in its own right. Because of this, current econometric research often involves the simultaneous analysis of multiple observed spells of the same type of duration for a given individual, or multiple observed spells of different types of durations for a given individual. For example, it may involve simultaneous and consecutive durations in labor market states and marital states. It may also involve the analysis of treatment effects on a duration variable, if the duration until treatment (or the duration of the treatment) is stochastic. In this chapter we therefore pay special attention to the analysis of multiple durations. We examine different types of relations between duration variables, as motivated by economic theory. We then examine the way in which they can be incorporated in multivariate extensions[1] of the MPH model, and we discuss identification of the determinants of these multivariate models as well as identification of deep structural parameters. For the case where the dependence runs by way of related unobserved explanatory variables (in which case we call the model a multivariate MPH (MMPH) model), we compare different parametric heterogeneity distributions. One of the main conclusions of the sections on multiple-duration models is that, in microeconometric research involving self-selection, duration data are much more informative than binary data. This is important because economic theory generally predicts the absence of exclusion restrictions based on characteristics of the individual under consideration, so that these can not be used for identification.

So far, we have been vague on the meaning of notions like "state", "duration", "exit rate", and "explanatory variable". In Section 2 we provide some formal definitions. We stress that the economic meaning of these notions is entirely context-dependent: what distinguishes states or transitions in one study may not be relevant in another study. Throughout the chapter we will be concerned with the economic insights that can be obtained from duration analysis. For that reason we outline in Section 3 some motivating underlying economic models for durations. In particular, we examine search models of individual labor market behavior. After these preparatory sections we examine the MPH model in Sections 4 and 5. Section 6 deals with the identification of the MPH model in case the data provide durations of multiple spells in a given state for a given individual. Such data are called multi-spell data. Again, the meaning of these notions is rather vague at this stage. Basically, the idea is that the data provide multiple

[1] In this chapter, "multivariate" refers to multiple durations and not to multiple explanatory variables.

independent drawings from the individual-specific duration distribution. Sections 7–9 deal with multiple-duration models in general. These constitute a very broad class of models, and they include, as a special case, the model of Section 6 with durations of multiple spells in a given state for a given individual. Section 10 concludes and provides recommendations on empirical approaches.

Throughout the chapter, time is taken to be continuous.[2] When specifying a duration distribution, the point of departure will invariably be the exit rate or hazard rate (this is motivated in Section 2). This implies that we do not focus on so-called Accelerated Failure Time models (see e.g. Kalbfleisch and Prentice, 1980), which enjoy some popularity outside economics. At times, though, we compare the latter models to models that are based on a specification of the hazard rate.

In this chapter we do not focus on estimation methods or specification tests. Applied studies generally use well-established estimation methods like Maximum Likelihood, Cox Partial Likelihood, Conditional Likelihood, or nonparametric methods. The book by Lancaster (1990), which is the most comprehensive volume on econometric duration analysis so far, provides an excellent survey on estimation methods and specification tests for MPH models in econometrics. Andersen et al. (1993) survey the literature on the modern statistical foundations. Kiefer (1988) and Yamaguchi (1991) lucidly explain the basics of the empirical analysis of duration models. Finally, the survey by Neumann (1997) discusses specification tests as well, and also pays attention to the estimation of structural (search) models.

2 Basic concepts and notation

Consider the spells experienced by certain subjects in a certain state. The duration of the spell is stochastic and is denoted by T, and realizations of T are denoted by t.[3] The cumulative distribution function of T is denoted by F, so

[2] See Meyer, 1995, for a survey of discrete-time reduced-form duration models. These models include continuous-time models where time is aggregated into intervals of unit length, as well as models where time is genuinely discrete.

[3] Throughout most of the chapter, we use t to denote the random variable as well as its realization. This abusive notation has become common in duration analysis because it allows for concise formulations that are generally unambiguous.

$F(t) = \Pr(T \leq t)$, with $F(0) = 0$. The *survivor function* of T is defined as one minus the distribution function and is denoted by $\overline{F}$, so

$$\overline{F}(t) = 1 - F(t)$$

As noted in the introduction, we restrict attention to continuous random variables T, and we denote a probability density function of T by f. In fact, $F, \overline{F}$, and f will be used as generic symbols for cumulative distribution functions, survivor functions, and probability density functions, respectively, and their arguments make clear which random variable is considered.

In a discrete-time setting, the *hazard function* of T at t is defined as the probability that the spell is completed at t given that it has not been completed before t, as a function of t. With T continuous, we define the hazard function as

$$\theta(t) = \lim_{dt \downarrow 0} \frac{\Pr(T \in [t, t + dt) | T \geq t)}{dt}$$

So, somewhat loosely, the hazard function is the rate at which the spell is completed at t given that it has not been completed before, as a function of t. The value of the hazard function (for a particular t, or for arbitrary t) is called the "hazard rate" or simply "the hazard". It is also called the "exit rate" to stress the fact that completion of the spell is equivalent to exit out of the state of interest. Again, we use θ as a generic symbol for a hazard, and its argument makes clear which random variable is considered. The hazard function $\theta(t)$ is said to be duration dependent if its value changes over t. Positive (negative) duration dependence means that $\theta(t)$ increases (decreases).

The hazard function provides a full characterization of the distribution of T, just like the distribution function, the survivor function, and the density function. All of these can be expressed in terms of one another. For $F, \overline{F}$, and f this is well known. Concerning θ, the following relations (which are easy to derive) express θ in terms of the other functions, and vice versa,

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

$$\overline{F}(t) = \exp\left(-\int_0^t \theta(u) du\right) \qquad t \geq 0 \tag{1}$$

The hazard function is the focal point of econometric duration models. That is, properties of the distribution of T are generally discussed in terms of properties of θ. There are two major reasons for this. First, and most importantly, this approach is dictated by economic theory. In general, theories that aim at explaining durations focus on the rate at which the subject leaves the state at duration t given that he has not done so yet. In particular, they explain the hazard at t in terms of external conditions at t as well as the underlying economic behavior of the subjects that are still in the state at t. Theoretical predictions about a duration distribution thus run by way of the hazard of that distribution. It is obvious that if the completion of a spell is at least partly affected by external conditions that change over time (e.g. due to external shocks), and if one attempts to describe behavior of the subject over time in a changing environment, then it is easier to think about the rate of leaving at t given that one has not done so than to focus on the unconditional rate of leaving at t. In the next section we provide some examples of such theories.

It is often stated that a major advantage of using the hazard function as a basic building block of the model is that it facilitates the inclusion of time-varying covariates. This is, of course, part of the argument of the previous paragraph; it reformulates the issue from the point of view of a builder of reduced-form models.

The second major advantage of using the hazard function as the basic building block of the model is entirely practical. Real-life duration data are often subject to censoring of high durations. In that case it does not make sense to model the duration distribution for those high durations.

Whereas the hazard function is the focal point of model building in duration analysis, the mean of the endogenous variable is the focal point in regression analysis. On some occasions in the chapter we compare duration models to regression models. For future reference it is useful to present the equation below. This equation follows directly from the fundamental result that the integrated hazard function $\int_0^t \theta(u)du$ has an exponential distribution[4] with parameter 1.

$$\log \int_0^t \theta(u)du = \varepsilon \tag{3}$$

[4] *Family of exponential distributions:*

$$f(t) = \vartheta e^{-\vartheta t} \qquad \text{for all } t \geq 0, \text{ with } \vartheta > 0 \tag{2}$$

Here, ε has an Extreme Value – Type I (EV1) distribution. This distribution does not have any unknown parameters; its density equals

Extreme Value – Type 1 distribution:

$$f(\varepsilon) = e^{\varepsilon} \cdot e^{-\exp(\varepsilon)}, \qquad \text{for all } -\infty < \varepsilon < \infty.$$

Equation (3) therefore again shows that once the hazard function is completely specified, then so is the duration distribution. Note that the transformation of t on the left-hand side of (3) can be interpreted as a particular change in the time measurement scale. The equation states that after this transformation, the only variation left in the duration concerns the purely random variation that is unrelated to the determinants of $\theta(t)$. Note that if one specifies a model for $\theta(t)$ then a natural model specification test follows from a comparison of the empirical distribution of the estimated left-hand side of (3) to the distribution of ε (see Lancaster, 1990).

3 Some structural models of durations

In this section we briefly discuss some economic-theoretical models that predict distributions of duration variables. These theoretical models have been structurally estimated using data on such duration variables, and they have been used to interpret estimates of reduced-form duration models. The common feature of the models is that they are search models, which describe the duration until an event as the outcome of a decision on the optimal moment of stopping the search for something desirable.[5] For expositional reasons we phrase the models in terms of search for jobs by individual agents on the labor market (although they are applicable to many other types of search). Job search models have been very popular as explanatory theoretical frameworks for reduced-form econometric duration analyses (see Devine and Kiefer, 1991).

[5] There are many other theoretical models that give rise to duration distributions. Examples are learning models (see e.g. Jovanovic, 1984) and dynamic discrete choice models (see e.g. Rust, 1994, for a survey). The latter can be considered as generalizations of basic search models although they are necessarily in discrete time; as such they give rise to discrete duration distributions. These models may also be used to explain multiple durations for a given subject (see e.g. Van der Klaauw, 1996).

3.1 Standard search model

3.1.1 Stationarity

In this subsection we consider the prototype job search model for the behavior of unemployed workers. Here, the duration variable of interest is the unemployment duration. Since this model has been discussed extensively many times (e.g. Mortensen, 1986), the present exposition is brief.

The model aims to describe the behavior of unemployed individuals in a dynamic and uncertain environment. Job offers arrive at random intervals following a Poisson process with arrival rate λ. A job offer is a random drawing (without recall) from a wage offer distribution with distribution function $F(w)$.[6] It is assumed that all jobs are full-time jobs. Every time an offer arrives, the decision has to be made whether to accept the offer or reject it and search further. Once a job is accepted it will be held forever at the same wage, so job-to-job transitions are excluded. It is assumed that individuals know λ and F but that they do not know in advance when job offers arrive and what wages are associated with them. During the spell of unemployment a benefit b is received. Unemployed individuals aim at maximization of their own expected present value of income over an infinite horizon. The subjective rate of discount is denoted by ρ.

The variables λ, w, b and ρ are measured per unit time period. It is assumed that the model is stationary. This means that λ, F, b and ρ are assumed to be constant, and, in particular, independent of unemployment duration and calendar time and independent of all events during unemployment. To ensure that attention is restricted to economically meaningful cases, and to guarantee the existence of the optimal strategy, we assume that $0 < \lambda, \mathrm{E}_F(w), b, \rho < \infty$. For ease of exposition we take F to be continuous.

Let R denote the expected present value of search when following the optimal strategy. Because of the stationarity assumption and the infinite-horizon assumption, the unemployed individual's perception of the future is independent of time or unemployment duration, so the optimal strategy is constant during the spell of unemployment and R does not depend on the elapsed unemployment duration t. It is well known (see e.g. Mortensen, 1986) that there is a unique solution to

[6] Note that F here denotes a distribution of wage offers rather than a duration distribution.

the Bellman equation for R, satisfying

$$\rho R = b + \lambda \mathrm{E}_w \max\{0, \frac{w}{\rho} - R\} \tag{4}$$

In this equation, the expectation is taken over the wage offer distribution F. Equation (4) has a familiar structure (see e.g. Pissarides, 1990). The return of the asset R in a small interval around t equals the sum of the instantaneous utility flow in this interval, and the expected excess value of finding a job in this interval. When an offer of w arrives at t then there are two options: *(i)* to reject it (excess value zero), and *(ii)* to accept it (excess value $w/\rho - R$). It is clear that the optimal policy is to choose option *(ii)* iff $w > \rho R$. Therefore, the optimal strategy of the worker can be characterized by a reservation wage ϕ: a job offer is acceptable iff its wage exceeds ϕ, with $\phi = \rho R$. Using equation (4), ϕ can be expressed in terms of the model determinants,

$$\phi = b + \frac{\lambda}{\rho} \int_{\phi}^{\infty} \overline{F}(w) dw$$

Note that this equation has a unique solution for ϕ.

The hazard (or exit rate out of unemployment, or transition rate from unemployment into employment) θ equals the product of the job offer arrival rate and the conditional probability of accepting a job offer,

$$\theta = \lambda \overline{F}(\phi)$$

As a result of the stationarity assumption, θ does not depend on the elapsed duration of unemployment. Consequently, the duration of unemployment t has an exponential distribution (see (2)) with parameter θ.

Versions of this model have been structurally estimated with individual data on unemployment durations and wages. “Structural” here means that the theoretical framework is assumed to describe the empirical distribution of durations and wages. This enables estimation of the determinants $\lambda, F, \ldots$ of individual behavior. See Yoon (1981), Flinn and Heckman (1982a), Narendranathan and Nickell (1985) and Van den Berg (1990b) for examples of this, and Wolpin (1995) for a survey.

3.1.2 Nonstationarity without anticipation

The stationarity assumption made in the previous subsection is often unrealistic. The values of the structural determinants may change because of duration dependence of the amount of unemployment benefits, a stigma effect of being long-term unemployed, policy changes, or business cycle effects. Sooner or later these features of the labor market and personal characteristics of job searchers are recognized and used in determining the optimal strategy. So, generally, the optimal strategy is not constant in case of nonstationarity.

To proceed, assume that the individual's search environment is subject to unanticipated changes in the values of the structural determinants. Thus, the values of these determinants may change over the duration, but the individual always thinks that they will remain constant at their current values. This might be a reasonable assumption in case of a change in λ that is due to a random macroeconomic shock, or in case of a change in b that is due to a sudden change in the benefits system.

By exploiting the analogy to the stationary model, we obtain the following equations for the reservation wage function $\phi(t)$, giving the reservation wage at time t, and the hazard function $\theta(t)$,

$$\phi(t) = b(t) + \frac{\lambda(t)}{\rho(t)} \int_{\phi(t)}^{\infty} \overline{F}(w|t)dw$$

$$\theta(t) = \lambda(t)\overline{F}(\phi(t)|t)$$

where $F(w|t)$ denotes the wage offer distribution at time t (so it should not be interpreted as a distribution conditional on the realization of a random duration variable). In general, $\theta(t)$ varies with t. The distribution function for the duration of unemployment subsequently follows from equation (1). See Narendranathan (1993) for a structural empirical analysis of a nonstationary model without anticipation.

3.1.3 Nonstationarity with anticipation

In many cases it is not realistic to assume that individuals do not anticipate changes in the values of λ, F, and b. In this subsection we consider nonstationarity with anticipation, along the lines of Van den Berg (1990a).[7] The structural

[7]Some special cases of this model have been examined earlier; see e.g. Mortensen (1986).

determinants λ, F, and b are allowed to vary over the duration t in a deterministic way (so dependence on past offer arrival times or wage levels associated with rejected offers is ruled out). This entails that the process with which job offers arrive is a non-homogeneous Poisson process. We assume that job searchers have perfect foresight in the sense that they correctly anticipate changes in the values of λ, F, and b. In other words, we expect people to know how these are related to t. As usual, individuals do not know in advance when job offers arrive, or which w are associated with them. Finally, we assume that λ, F, and b are constant for all sufficiently high t. The latter implies that the optimal strategy is also constant for sufficiently high t.

Let $R(t)$ denote the expected present value of search if unemployment duration equals t, when following the optimal strategy. Under regularity conditions, there is a unique continuous solution to the Bellman equation for $R(t)$, satisfying

$$\rho R(t) = \frac{dR(t)}{dt} + b(t) + \lambda(t).\mathrm{E}_{w|t} \max\{0, \frac{w}{\rho} - R(t)\}$$

at points at which $R(t)$ is differentiable in t, where the expectation is taken over the wage offer distribution $F(w|t)$ at t. Notice the similarity with equation (4) above. The return of the asset $R(t)$ in a small interval around t equals the sum of the appreciation of the asset in this interval, the instantaneous utility flow in this interval, and the expected excess value of finding a job in this interval. The optimal strategy can be characterized by a reservation wage function $\phi(t)$ that gives the reservation wage at time t. Using the fact that $\phi(t) = \rho R(t)$, it follows that

$$\frac{d\phi(t)}{dt} = \rho\phi(t) - \rho b(t) - \lambda(t) \int_{\phi(t)}^{\infty} (w - \phi(t)) dF(w|t)$$

This differential equation has a unique solution for $\phi(t)$, given the boundary condition that follows from the assumption that the model is stationary for all sufficiently high t.

The hazard function $\theta(t)$ now equals

$$\theta(t) = \lambda(t)\overline{F}(\phi(t)|t)$$

In general, $\theta(t)$ varies with t. The distribution function for the duration of unemployment subsequently follows from equation (1).

For examples of structural empirical analyses of nonstationary models with anticipation, see Wolpin (1987), Van den Berg (1990a), Engberg (1991), and Garcia-Perez (1998).

3.2 Repeated-search model

Models of repeated search allow the economic agent to search further for better matches after a match has been formed. The best-known model of repeated search is the so-called on-the-job search model which aims to describe the behavior of employed individuals who search for a better job (see Mortensen, 1986, for an overview). In the basic on-the-job search model, a job is characterized by its wage w which is taken to be constant within a job. For a working individual, the search environment is specified in exactly the same way as we did in Subsection 3.1.1 for an unemployed individual. In particular, we assume the model to be stationary. The optimal strategy is constant during a job spell, and the the expected present value of search $R(w)$ when following the optimal strategy in a job with wage w satisfies

$$\rho R(w) = w + \lambda \mathrm{E}_{w^*} \max\{0, R(w^*) - R(w)\}$$

where the expectation is taken with respect to the distribution F of wage offers w^*. Clearly, the optimal strategy is such that one accepts a job if and only if the offered wage w^* exceeds the current wage w, so it suffices to compare instantaneous income flows (i.e., the optimal strategy is "myopic"), and the reservation wage simply equals the current wage.

For a given current wage w, the hazard of the job duration distribution (or exit rate out of the present job) equals

$$\theta = \lambda \overline{F}(w)$$

As a result, the duration of a job with a wage w has an exponential distribution with this parameter θ. Note that models of repeated search are informative on the joint distribution of consecutive job durations.

If, during employment, exogenous separations occur at a rate δ, then this does not affect the optimal strategy. The exit rate out of the present job then equals

$\lambda\overline{F}(w)+\delta$. See Flinn (1996) for an example of structural estimation of this model with job duration data.[8]

Burgess (1989) introduces a rather manageable type of nonstationarity in this model. The individual's search environment (i.e., λ and F) is subject to shocks that are not job-specific but rather such that they act similarly on all employed workers. The shocks may be anticipated or unanticipated. It is intuitively obvious that this nonstationarity does not change the optimal strategy: it remains optimal to accept another job if and only if its wage exceeds the current wage. We thus obtain for the job-to-job transition rate,

$$\theta(t) = \lambda(t)\overline{F}(w|t)$$

Throughout the remainder of the chapter, it is important to keep in mind that empirical duration analysis is ultimately interested in structural parameters that represent determinants of individual behavior. This is also true for empirical analysis in which reduced-form models are estimated that are not explicitly specified as a theoretical model. In the sequel we return to this issue.

4 The Mixed Proportional Hazard model

4.1 Definition

For sake of convenience, we use the term "individual" to denote the subject that experiences certain spells in a given state. We consider the population of individuals that consists of the inflow into this given state. This can be the inflow at a given point of time, or the inflow at any time. We assume that, for a given individual in this population, the subsequent duration T is an absolutely continuous and positive random (duration) variable. The distribution of T (or, equivalently, the hazard function) may vary across individuals. We assume that all individual variation in the hazard function can be characterized by a finite-dimensional vector of observed explanatory variables (or "covariates", or "regressors") x and an

[8] The empirical analysis of so-called equilibrium search models, which endogenize the wage offer distribution F, often involves the joint estimation of the distributions of unemployment durations and job durations. See e.g. Van den Berg and Ridder (1998), Bontemps, Robin and Van den Berg (2000), and Bowlus, Kiefer and Neumann (2001).

unobserved heterogeneity term v. The latter term can be interpreted as a function of unobserved explanatory variables.[9] In this subsection we assume that x is time-invariant, and consequently we define the Mixed Proportional Hazard model as a model with time-invariant explanatory variables. In the next subsection we introduce time-varying explanatory variables.

For an individual with explanatory variables x and unobserved heterogeneity v, the hazard function of the random variable T evaluated at the duration t is denoted by $\theta(t|x,v)$. This notation highlights the fact that we condition on x and v. The standard MPH model is now defined by

Definition 1 : Standard MPH model. *There are functions ψ and θ_0 such that for every t and every x and v there holds that*

$$\theta(t|x,v) = \psi(t) \cdot \theta_0(x) \cdot v \tag{5}$$

This model was developed by Lancaster (1979), which includes an empirical application to unemployment duration data, and by Vaupel, Manton and Stallard (1979).[10] The function $\psi(t)$ is called the "baseline hazard" since it gives the shape of the hazard function for any given individual. Only the *level* of the hazard function is allowed to differ across individuals. The term $\theta_0(x)$ is called the "systematic part" of the hazard. In applied work, it is common to specify

$$\theta_0(x) = \exp(x'\beta), \tag{6}$$

so that $\theta(t|x,v)$ is multiplicative in all separate elements of x.

For convenience, we make a number of regularity assumptions on the determinants of the model.

Assumption 1 *The vector x is k-dimensional with $1 \leq k < \infty$. The function $\theta_0(x) : \mathcal{X} \subset \mathbb{R}^k$ is positive for every $x \in \mathcal{X}$.*

Assumption 2 *The function $\psi(t)$ is positive and continuous on $[0, \infty)$, except that $\lim_{t \downarrow 0} \psi(t)$ may be infinite. For every $t \geq 0$ there holds that $\int_0^t \psi(\tau) d\tau < \infty$, while $\lim_{t \to \infty} \int_0^t \psi(\tau) d\tau = \infty$.*

[9] Lancaster (1990) shows that v to some extent may also represent measurement errors in T and x.

[10] Nickell (1979) contains the first estimation of a discrete-time MPH-type model.

Assumption 3 *The distribution G of v in the inflow satisfies* $\Pr(0 < v < \infty) = 1$.

Assumption 4 *The individual value of v is time-invariant.*

It should be stressed that, for virtually all of the results in the chapter, these conditions are stronger than needed. This is particularly true for Assumption 2. It is often sufficient that $\psi(t)$ is integrable, and sometimes it is sufficient that $\int_0^t \psi(\tau)d\tau < \infty$ only on some interval. For expositional reasons, we do not deal with this. On the other hand, for identification, additional assumptions are needed (see Section 5). We do not list those here because it is interesting to contrast alternative assumptions in the light of identifiability issues.

It is useful to examine the special case in which there is no unobserved heterogeneity ($v \equiv 1$). In that case the model is called a Proportional Hazard (PH) model (this model was developed by Cox, 1972, and predates the MPH model). The PH model specification is regarded to be simple and yet sufficiently rich to capture many data properties. The popularity of the PH model in reduced-form duration analysis is comparable to the popularity of the linear regression model in reduced-form regression analysis. Note that the general regression-type expression for the integrated hazard function (see (3)) reduces to

$$\log \int_0^t \psi(u)du = -x'\beta + \varepsilon \qquad (7)$$

for the PH model, where we substituted (6) and ε has an EV1 distribution. It should again be stressed that ε represents the purely random variation in the duration outcome – it does not capture unobserved individual characteristics. In comparison to a linear regression model (say $\log t = x'\beta + \varepsilon$, with ε having an unknown distribution with mean zero), the left-hand side of (7) has a more general specification, since it involves an unknown transformation of the duration variable, whereas the right-hand side has a more restrictive specification, since the distribution of the error term is completely specified. Thus, the PH model and the regression model are not nested, and they derive their flexibility from different sources.

The β parameters in the linear regression model are estimated consistently by OLS under a wide range of distributions of ε. Similarly, the β parameters in the PH model are estimated consistently by Partial Likelihood under a wide range of specifications of the baseline hazard $\psi(t)$. More precisely, the β parameters are estimated consistently by maximization of a partial likelihood function that does not depend on the baseline hazard function, which can be estimated nonparametrically in a second stage (see Lancaster, 1990, for details). This is arguably one of the great advantages of the PH model, but it does not carry over to the MPH model in general.

For the MPH model, equation (3) reduces to

$$\log \int_0^t \psi(u)du = -x'\beta - \log v + \varepsilon \tag{8}$$

where again we substituted (6), and where again ε has an EV1 distribution. The equation states that the log integrated baseline hazard function given x has the same distribution as the distribution of a random variable that is the sum of an EV1 random variable and another random variable (namely $-x'\beta - \log v$ given x). Since we have not made an assumption on the distribution of v, it is clear that specification (8) is much more general than (7). Now we have a flexible specification for both the transformation of t and the distribution of the error term. However, the latter distribution cannot be just any distribution. For example, it cannot be a normal distribution, because the sum of an EV1 random variable and another random variable cannot have a normal distribution (see Ridder, 1990). It turns out that the MPH model is actually identified under an assumption on the tail of the distribution of v (see Section 5).

We end this subsection by mentioning some other reduced-form duration models. Consider the following model,

$$\log z(t) = -x'\beta + \epsilon \tag{9}$$

with $z(t)$ positive and increasing in t. This reduces to the MPH model if the "error term" ϵ is distributed as the sum of an EV1 random variable and another random variable. If no assumption is made on the distribution of ϵ then (9) is called a "transformation model" (see Horowitz, 1996). If it is subsequently imposed that $z(t) = t$ then we obtain the Accelerated Failure Time (AFT) model,

$$\log t = -x'\beta + \epsilon$$

For future reference it is useful to note that in the AFT model the survivor function can be written as

$$\overline{F}(t|x) = \exp(-\Psi(t \cdot e^{x'\beta})) \tag{10}$$

where Ψ is the integrated hazard function of the random variable $\exp(\epsilon)$. Clearly, the individual characteristics act on the duration distribution by transforming the time scale from t to $t\exp(x'\beta)$. This may be an accurate description of the actual variation in the lifetime distributions of complex self-evolving organisms or mechanisms. Because of the one-to-one relation between a distribution and its hazard function, the AFT specification can be translated into a specification of the hazard function of $t|x$. Obviously, the latter need not be an MPH specification. Note that in the transformation model and the AFT model, the hazard does not serve as the focal point of model specification. This has strongly limited the use of these models in social science duration analyses. We return to this in Subsection 5.6.

4.2 Time-varying explanatory variables

In practice, explanatory variables are often time-varying, and there are often good reasons to assume that the hazard function is affected by the current value of the explanatory variable (instead of e.g. its value at the beginning of the spell). In this subsection we discuss the incorporation of such explanatory variables in the PH model and (at the end of the subsection) the MPH model. Given that the chapter avoids measure theory, the exposition in this subsection is restricted to be rather informal, and we refer the reader to the references below for more rigorous analyses.

At first sight it may seem that time-varying explanatory variables can be incorporated in the PH model by replacing x by $x(t)$,

$$\lim_{dt \downarrow 0} \frac{\Pr(T \in [t, t+dt) | T \geq t, \{x(u)\}_0^t)}{dt} = \psi(t) \cdot \theta_0(x(t)) \tag{11}$$

where $\{x(u)\}_0^t$ denotes the time path of x up to t, and where, possibly, $\theta_0(x(t)) = \exp(x(t)'\beta)$. However, there are some caveats here. First, the values of the explanatory variables at t may in some sense be endogenous. The subject under study may have inside information at t on the future realization of the random variable T, and this information may affect the values of his observed explanatory variables at t and his hazard rate at t. It may then be erroneously concluded that the observed explanatory variables have a causal effect on the duration. Consider an unemployed individual who knows that he will start to work in a job at a given future date and may for that reason decide not to enrol in a training program at t. If this is ignored in the empirical analysis then the effect of the number $x(t)$ of completed training programs at t on the exit rate out of unemployment at t may be under-estimated. A second caveat concerns the fact that $x(t)$ could cause the duration distribution to be discontinuous at certain durations. This would complicate the statistical and empirical analysis.

To proceed, assume that the time-varying explanatory variables constitute a stochastic process $X = \{X(t) \, : \, t \geq 0\}$. Without loss of generality we take $X(t)$ to represent *all* explanatory variables for the hazard rate at t. Note that we may trivially include time-invariant or fully deterministic explanatory variables in X, and recall that for the time being we assume that all heterogeneity is observed. Kalbfleisch and Prentice (1980) develop a classification of duration models with time-varying covariates, in order to describe classes for which standard econometric procedures can be applied. This classification is rather vague and not exhaustive (Heckman and Taber, 1994). Fortunately, the recent mathematical-statistical literature on counting processes and martingales has allowed a breakthrough on these issues. The counting process approach assumes that the durations, the values of the time-varying explanatory variables, and the observational plan, are all outcomes of stochastic processes (as such, it allows for quite general censoring schemes; see Fleming and Harrington, 1991, Andersen and Borgan, 1985, and Andersen et al., 1993, for excellent surveys, and Ridder and Tunalı, 1999, for an exposition which also avoids measure theory and includes an econometric application). The approach focuses on a PH model framework in which X has the property that:

- X is a predictable process.

Here, predictability basically means that the values of all explanatory variables for the hazard at t must be known (and observable to the researcher) just before t. In other words, the values of the variables which capture all individual variation in the hazard rate at t must be known and observable at t^-. In yet other words, the values of the explanatory variables at t are influenced only by events that have occurred up to time t, and these events are observable. The information on the values *at* time t does not help in predicting a transition at t. Note that predictability does *not* mean that the whole future realization of X can be predicted at some point in time. Below we give some examples. As Ridder and Tunalı (1999) point out, the concept of predictability is basically the same as the concept of weak exogeneity in time series analysis (and is thus weaker than the concept of strong exogeneity). In addition to predictability, we need a technical assumption which basically ensures that the realized outcomes of $X(t)$ and $\theta_0(X(t))$ are bounded. Fleming and Harrington (1991) contains a more precise exposition with explicit use of measure theory. The counting process approach has been very successful in the derivation of (asymptotic) properties of estimators and test statistics for general settings, including generalizations of the commonly used estimators and test statistics in duration analysis (see the references above).

Now consider the stochastic process $\Pr(T \leq t|\{X(u)\}_0^t)$, which is a process given the evolution of X up to t, as a function of t. Assume that this process is absolutely continuous. Sufficient for this (in addition to the predictability of X) is, basically, that T does not have a strictly positive probability of occurrence at t, given X up to t. Given absolute continuity, the counting process model can be expressed as a model of hazard functions. Conversely, a PH model of hazard functions, with X having the above properties, and with absolute continuity of the above process, can be thought of as being generated by a PH counting process model (Fleming and Harrington, 1991, Arjas, 1989). It should be noted that these results have been derived for models with

$$\theta_0(X(t)) = \exp(X(t)'\beta)$$

and certain other specifications of θ_0 (see Andersen and Borgan, 1985).

The results imply that if we start off with a PH-type model of a hazard function, and X has the properties above, then we can perform valid econometric inference using standard methods, on the basis of specification (11) for the hazard rate. This is, in a nutshell, why predictability of the time-varying explanatory variable is an extremely useful property. Given predictability, we may apply the standard tools of duration analysis.[11]

It is useful to examine the predictability for some special cases for X. First, if X is time-invariant then it is obvious that it is predictable. Now suppose its path is fully known in advance. For example, the unemployment benefits level

[11] Note that in case of stochastic explanatory variables it does not make sense to talk about "the" probability distribution of T.

as a function of the elapsed unemployment duration may be determined at the date of inflow into unemployment, by the institutional setting. Clearly, X is then predictable as well. If X is stochastic then somewhat loosely one may state that if the current value of X only depends on past and outside random variation then X is predictable (Andersen and Borgan, 1985). Now consider the case in which the individual has inside information on future realizations of X. For example, an unemployed individual may expect a baby or may expect participation in a training program at a future date. This information may be used as input in the individual's decision problem and as a result may affect the current hazard rate. If this information is not known to the analyst then X is not predictable. The same is true if the individual anticipates the realization of T and if this affects the current hazard. Note that it is intuitively plausible that, in these cases, standard inference may lead to inconsistent estimates. These cases include so-called instantaneous feedback effects: predictability is not satisfied if X jumps in an unexpected way at t. This does not mean that jumps in regressor values are not allowed at all if one demands predictability. Suppose that one wants to model that an individual's hazard rate increases by a certain amount immediately after the realization of another duration variable τ which is independently distributed from the duration of interest and from other time-varying covariates. This can be captured by a time-varying regressor $\mathrm{I}(t > \tau)$, which is predictable.

Now consider the case where a time-invariant explanatory variable is unobserved (i.e., consider MPH models). If we condition on the unobserved heterogeneity value v and do as if v is observed then the above analysis remains valid. If v is treated as unobserved then v is not predictable. As we shall see in Section 5, ignoring the unobserved heterogeneity in empirical inference generally leads to inconsistent inference. In this case, the standard solution is to jointly model the hazard function and the distribution of v, and to integrate v out of the likelihood.

We end this subsection by making a few comments. First, time-varying explanatory variables may play a very different role in *other* reduced-form duration models, such as the AFT model. This reflects the fact that such models do not take the hazard function as the point of departure for the model specification.[12] Secondly, as noted above, the counting process approach allows for quite general censoring schemes; in fact, what is needed is that the observational plan is a predictable process. Thirdly, in the remainder of the chapter, the focus is mostly on models without time-varying explanatory variables. The motivation for this is basically the same as the one (implicitly) adopted in most of the methodological literature on duration models, namely that the analysis of these models is relatively manageable and that the results create a good starting point for future analysis of more general models. Below, whenever we encounter time-varying explanatory variables, we tacitly assume that the conditions that ensure valid inference with standard methods are satisfied.

4.3 Theoretical justification

As mentioned above, the MPH model and its special cases are often regarded to be useful reduced-form models for duration analysis. The resulting estimates are generally interpreted with the help of some economic theory. However, the MPH model specification is not derived from economic theory, and it remains to be seen whether the MPH specification is actually able to capture important theoretical relations, and, conversely, whether the MPH specification can be generated by theory.

The main assumption underlying the MPH model is that the three determinants of the hazard act multiplicatively on the hazard. This implies that if the elapsed duration has a positive effect on the hazard, then this effect is stronger for individuals with characteristics that also have a positive effect on the hazard. Of course, the distinction between two of the three determinants (the observed and unobserved explanatory variables) is only relevant from an empirical point of view. If the researcher could observe all determinants without measurement error, then the unobserved heterogeneity term can be omitted. Within a theoretical

[12] For example, consider the formulation (10) of the AFT model. Typically, time-varying explanatory variables are included in this model by way of

$$\overline{F}(t|\{X(u)\}_0^t) = \exp(-\Psi(\int_0^t \exp(X(u)'\beta)du))$$

In that case, the hazard rate at t depends on the whole history $\{X(u)\}_0^t$ of X.

framework it is irrelevant whether a certain background variable can be observed by the researcher or not. This means that from a theoretical point of view, the most important assumption of the MPH model is that the elapsed duration and the explanatory variables act multiplicatively on the hazard.

In economics, this assumption is often hard to justify. We illustrate this by examining the economic theories discussed in Section 3.[13] First consider the job search model of Subsection 3.1.2. We allow all structural determinants to differ across individuals, and this is captured by time-invariant explanatory variables x. We assume that the analyst observes x (and the duration t) but does not directly observe how the structural determinants, the optimal strategy, or the acceptance probability change with t. If such changes would be directly observed then obviously it would make sense to include them as time-varying explanatory variables. We return to time-varying explanatory variables towards the end of the subsection.

From Subsection 3.1.2 we obtain the following system of equations, in obvious notation,

$$\phi(t,x) = b(t,x) + \frac{\lambda(t,x)}{\rho(t,x)} \int_{\phi(t,x)}^{\infty} \overline{F}(w|t,x)dw$$

$$\theta(t,x) = \lambda(t,x)\overline{F}(\phi(t,x)|t,x)$$

Intuitively, the main reason for why it is difficult to obtain a multiplicative structure for $\theta(t,x)$ is that in general $\overline{F}(\phi(t,x)|t,x)$ is not multiplicative in ϕ, which in turn depends on "everything in the model" in a non-multiplicative fashion. Below are a few special cases where the resulting $\theta(t,x)$ *is* proportional in t and x. Note that these assume that changes in the structural determinants are unanticipated.

Example 1. Let F be a Pareto distribution,

Family of Pareto distributions:

$$\overline{F}(w) = (w_0/w)^{\nu} \qquad \text{for all } w > w_0, \text{ with } w_0, \nu > 0 \tag{12}$$

[13] The problem is more general, though.

where we actually assume $\nu > 1$ to ensure that the optimal strategy exists, and where the parameters w_0 and ν of F may depend on t and x. Let in addition $b \equiv 0$. Then

$$\theta(t,x) = \rho(t,x)(\nu(t,x) - 1)$$

Let the discount rate ρ vary with x but not with t, and let the shape parameter ν vary with t but not with x (for example, long-term unemployed workers receive on average lower wage offers). Then the hazard is proportional in t and x. Of course, the same result applies if ρ only varies with t and ν only with x. Also, if ν is a fixed constant and ρ is proportional in t and x, then the hazard is proportional as well. Note that the assumption $b \equiv 0$ is very strong.[14]

Example 2. Let $\rho = \infty$, so that workers do not care about the future. Then $\phi \equiv b$, and

$$\theta(t,x) = \lambda(t,x)\overline{F}(b(t,x)|t,x)$$

If $\lambda(t,x)$ varies with t (e.g. because the long-term unemployed are stigmatized) but not with x, and F and b vary with x but not with t, then the hazard is proportional in t and x. Alternatively, if F and b do not depend on either t or x and λ is proportional in t and x, then the hazard is proportional as well.

Example 3. Let the structural determinants be such that ϕ is always smaller than the lowest wage in the market (e.g., benefits are so low that the reservation wage is below the mandatory minimum wage). Then $\overline{F}(\phi) = 1$ always, and

$$\theta(t,x) = \lambda(t,x)$$

so, if λ is proportional in t and x, then the hazard is proportional as well.

Example 4. This case is based on Yoon (1985), which is one of the very few studies to date on the theoretical justification of the PH model. He examines a model

[14] In general, if one is prepared to adopt a linearized specification for the reservation wage $\phi(t,x)$ as a function of its determinants, and if F has a Pareto distribution or an exponential distribution, then it is less difficult to obtain a multiplicative specification for $\theta(t,x)$. See Lancaster (1985a).

where jobs have a fixed and common tenure T^*, after which the individual dies.[15] The variable b is assumed to equal benefits minus search costs, and the model requires that the net value of b is negative. There is no discounting of the future (so the limiting case $\rho \downarrow 0$ is considered). It is straightforward to show that $\phi(t)$ then follows from

$$-b(t,x) = \lambda(t,x)T^* \int_{\phi(t,x)}^{\infty} \overline{F}(w|t,x)dw$$

Let F be a Pareto distribution (see (12)) with a fixed parameter $\nu > 1$ and a parameter $w_0(t,x)$. It follows that

$$\theta(t,x) = [\lambda(t,x)]^{\frac{-1}{\nu-1}} \, [w_0(t,x)]^{\frac{-\nu}{\nu-1}} \, \left[\frac{-b(t,x)(\nu-1)}{T^*}\right]^{\frac{\nu}{\nu-1}}$$

Obviously, there are many ways to obtain a PH specification from this.

Now consider anticipated changes in the structural determinants, i.e., consider the nonstationary job search model of Subsection 3.1.3. In particular, for ease of exposition, consider a special case where the only change concerns a drop in b at a duration τ (from b_1 to b_2). There still holds that $\theta(t,x) = \lambda(t,x)\overline{F}(\phi(t,x)|t,x)$. However, now the reservation wage $\phi(t)$ for $t < \tau$ depends on b_1 and b_2 as well as on $\tau - t$. The smaller the remaining time interval $\tau - t$ until the drop in b, the more important the future benefits level b_2 is for the current present value. As shown by Van den Berg (1990a, 1995), there are two reasons for this. First, the discounting of the future means that the far future carries less weight than the near future. Second, there is a probability that the individual leaves unemployment before τ, and this probability is lower if τ is in the near future. This probability depends on the hazard function itself, in between t and τ. As a result of all this, as the duration $t < \tau$ proceeds, the effect on the hazard of b_1 diminishes, and the effect of b_2 increases (with a magnitude that depends on all structural determinants). After τ, the hazard does not depend on b_1 anymore. It seems to be impossible to justify a PH specification with such a theoretical model, except for the following limiting case.

[15] Job separations leading to unemployment rather than death or permanent retirement are hard to reconcile with unanticipated duration dependence of the structural determinants, because of the repetitive nature of unemployment.

Example 5. Let $\rho \to \infty$ in the nonstationary job search model, so workers do not care about the future. In that case, even though an individual does have information on future changes, this does not affect his optimal strategy, and the exit rate out of unemployment is the same as in Example 2.

Finally, consider the nonstationary on-the-job search model of Subsection 3.2, and, in particular, the job-to-job transition rate (which will be our hazard rate). Note that there is no "feedback" from the structural determinants to the value of the reservation wage w. There holds that $\theta(t,x) = \lambda(t,x)\overline{F}(w|t,x)$, where x may include w, and the following result emerges.

Example 6. Let F be time-invariant in the nonstationary on-the-job search model. Then

$$\theta(t,x) = \lambda(t,x)\overline{F}(w|x)$$

which supports a PH specification if $\lambda(t,x)$ is multiplicative in t and x. If F has a Pareto distribution (see (12)), then its parameter w_0 is allowed to depend on t.[16,17]

The main conclusions of this subsection are as follows. First, the proportionality restriction of the (M)PH model can in general not be justified on economic-theoretical grounds. Second, if the optimal strategy is myopic (e.g. because of repeated search, or because the discount rate is infinite), then this restriction often follows from economic theory.

Despite the first conclusion, the (M)PH model has become very popular in reduced-form duration analysis, in particular in labor economics. The popularity of a reduced-form model that does not nest many structural models distinguishes duration analysis from the reduced-form analysis of wage data with the linear

[16] The proportionality results in Examples 4 and 6 can also be generated with other families of wage offer distributions than the Pareto family. Notably, F can be exponentially distributed, so $\overline{F}(w) = \exp(-\nu(w-w_0))$ on $w > w_0$, with $\nu > 0$.

[17] Here, as in previous examples, if the job offer arrival rate depends on an optimally chosen search intensity, then the scope for multiplicative specifications is further reduced. This is because this search intensity is a second "channel" through which all structural determinants affect the hazard in a non-multiplicative fashion (see e.g. Mortensen, 1986, for a theoretical analysis of such models).

regression model, since the linear specification has been justified extensively by human capital theory and traditional labor supply theory. Part of the attractiveness of the (M)PH model stems from the fact that it is difficult to think of a more parsimonious specification of the hazard that includes all single major determinants of it. (Also, recall that the Partial Likelihood estimation method allows for estimation of the systematic hazard of the PH model without the need to parameterize or estimate the baseline hazard.) In practice, the empirical application at hand does not always dictate a natural theoretical framework, and sometimes the scope of the application does not warrant a full-blown theoretical or structural analysis. In such cases, the (M)PH model is a useful framework whose properties have been thoroughly studied in the literature.

Last but not least, the MPH framework can be extended to a certain extent to incorporate some features of the theory at hand. Notably, changes over t in the value of a variable x can be incorporated by the inclusion of time-varying covariates. For example, in the study of unemployment insurance benefits on exit out of unemployment, the effect of the remaining benefit entitlement can be included as a time-varying covariate (see e.g. Solon, 1985). Also, if the data provide direct observations on how a structural determinant, the reservation wage, or the acceptance probability change over time, then these can be included as time-varying covariates. As an example, consider the models of Subsection 3.1, and suppose that $\phi(t,x)$ is fully observed and F is a time-invariant Pareto distribution which does not vary with x. Then $\theta(t,x) = \lambda(t,x)w_0^{\nu}[\phi(t,x)]^{-\nu}$, so if $\lambda(t,x)$ is multiplicative in t and x then this supports a PH specification with a time-varying covariate. As another example, consider the on-the-job model. One may observe business cycle indicators and use these as representations of $\lambda(t,x)$. Finally, changes in the effect over t of a variable x can be incorporated by the inclusion of interactions between t and x in the hazard.[18]

These extensions lead to less transparent models, and some of the distinct advantages of the MPH model are lost this way (see Section 5). Moreover, it should be stressed that the insertion of some time-varying covariates or time-varying parameters into an MPH model more often than not does not lead to a specification that can be generated by a theoretical model. This is intuitively

[18] One may use a nonparametric estimation method for an unrestricted specification of the hazard rate $\theta(t,x)$, allowing for full interactions. See e.g. Dabrowska (1987).

clear from the nonstationary model in which unemployment benefits decrease with the duration of unemployment.

As noted in Subsection 4.1, in applied work it is often assumed that each explanatory variable acts multiplicatively on the hazard rate (i.e., $\theta_0(x) = \exp(x'\beta)$). From the discussion above it is clear that economic theory often predicts that the different structural determinants do not act multiplicatively on the hazard. Thus, if each determinant is represented by different elements of x, then these elements interact with each other in the hazard. This can be incorporated to a certain extent in the MPH model, as inclusion of interaction terms for the different elements of x does not violate the (M)PH specification.[19]

We end this section by noting that the economic justification of other popular reduced-form duration model specifications is at least as difficult as the justification of the (M)PH specification. This holds in particular for the Accelerated Failure Time model, in which the mean of $\log t$ is specified as a linear function of x, so $\log t = -x'\beta + \epsilon$, and also for the additive hazard model, in which $\theta(t|x)$ is specified as $\theta(t|x) = \psi(t) + \theta_0(x)$. These two types of reduced-form duration models enjoy popularity in biostatistics, where the relation between theory and application is less compelling than in econometrics. Discrete-time reduced-form duration model specifications are also difficult to justify; they often do not follow from the underlying economic models (like discrete-time search models or dynamic discrete-choice models).

5 Identification of the MPH model with single-spell data

5.1 Some implications of the MPH model specification

In this section we examine identification of the MPH model with unobserved heterogeneity,[20] if the data provide i.i.d. drawings from the conditional distri-

[19]As an example, job search theory predicts that the elasticity of the exit rate out of unemployment with respect to unemployment benefits depends on the level of the benefits. This can be captured to some extent in a reduced-form analysis by including $(\log b)^2$ as an additional regressor (see Van den Berg, 1990c, for details).

[20]Identification of the determinants of the PH model is trivial if it is known that the data are generated by a PH model.

bution of $t|x$. In reality, the observations on t may be right-censored (i.e., for some observations it is only known that t exceeds a certain value) or interval-censored (e.g. if durations are grouped into intervals), or the sampling design may be non-random. Heckman and Singer (1984a), Ridder (1984), and Lancaster (1990) contain extensive examinations of the implied duration distributions in other sampling designs. Situations in which the data provide multiple durations for the same individual are discussed in subsequent sections.

Throughout the section we make the following model assumption,

Assumption 5 : Independence of observed and unobserved explanatory variables. *In the inflow, v is independent of x.*

Note that this assumption is stronger than the usual assumption in linear regression models that x and ε are uncorrelated or that they satisfy $\mathrm{E}(\varepsilon|x) = 0$.

It is useful to examine the distribution $F(t|x)$ of $t|x$ and derive the well-known result that the duration dependence of the hazard function $\theta(t|x)$ of this distribution is more negative than the duration dependence of the hazard function $\theta(t|x, v)$ (Lancaster, 1979, was the first point out these results; see also the survey in Lancaster, 1990, and Heckman and Singer, 1984a, who consider a generalization of the MPH framework).

By definition, we have

$$F(t|x) = \int_0^\infty F(t|x, v)\, dG(v) \tag{13}$$

where G is the cumulative distribution function of v in the inflow into the state of interest, and where $F(t|x, v)$ has the associated hazard function $\theta(t|x, v)$. Consequently, $\theta(t|x)$, which by definition equals $f(t|x)/\overline{F}(t|x)$, can be written as

$$\theta(t|x) = \frac{\int_0^\infty \theta(t|x, v)\, \overline{F}(t|x, v)\, dG(v)}{\overline{F}(t|x)} \tag{14}$$

By Bayes' Theorem, we have for every t that

$$dG(v|T > t, x) = \frac{\overline{F}(t|x, v)\, dG(v)}{\overline{F}(t|x)} \tag{15}$$

(Note that here we use T to denote a random variable.) In general, therefore, the distribution of $v|T > t, x$ depends on x for all $t > 0$, even though it does not for $t = 0$. The composition of the sample of survivors (as captured by the distribution of v) changes as time proceeds, in a way that that depends on t and x. This is an important aspect of the dynamic self-selection that occurs if one examines subsamples of individuals with higher and higher durations.

By substituting (15) into (14) we obtain $\theta(t|x) = E_{v|T>t,x}(\theta(t|x,v))$. Therefore,

$$\theta(t|x) = \psi(t) \cdot \theta_0(x) \cdot \mathrm{E}(v|T > t, x) \tag{16}$$

Let us denote the integrated baseline hazard at t as $z(t)$,

$$z(t) = \int_0^t \psi(\tau)\, d\tau$$

Of course, $-\log \overline{F}(t|x,v)$ equals $v \cdot \theta_0(x) \cdot z(t)$. By substituting this into equations (13) and (15) it follows that we can write

$$\mathrm{E}(v|T > t, x) = \frac{\int_0^\infty v \cdot e^{-v\cdot\theta_0(x)\cdot z(t)}\, dG(v)}{\int_0^\infty e^{-v\cdot\theta_0(x)\cdot z(t)}\, dG(v)} \tag{17}$$

It is useful to rewrite $\theta(t|x)$ in some different ways. First, note that the denominator on the right-hand side of (17) (which equals $\overline{F}(t|x)$) is nothing but the Laplace transform $\mathcal{L}$ of the distribution of v, evaluated at $\theta_0(x) \cdot z(t)$,

$$\mathcal{L}(s) = \int_0^\infty e^{-s\cdot v}\, dG(v) \tag{18}$$

Consequently, the numerator in (17) is nothing but minus the derivative of $\mathcal{L}$ evaluated at $\theta_0(x) \cdot z(t)$. This means that we can rewrite equation (16) as follows,

$$\theta(t|x) = \psi(t) \cdot \theta_0(x) \cdot \frac{-\mathcal{L}'(\theta_0(x) \cdot z(t))}{\mathcal{L}(\theta_0(x) \cdot z(t))} \tag{19}$$

So all derivatives of this with respect to x and/or t depend on G only by way of (derivatives of) the Laplace transform of G, evaluated at $\theta_0(x)z(t)$. Equivalently, all derivatives of $\theta(t|x)$ with respect to x and/or t depend on G by way of moments of $v|T > t, x$. Specifically,

$$\frac{d\log\theta(t|x)}{dt} = \frac{\psi'(t)}{\psi(t)} - \frac{\mathrm{Var}(v|T > t, x)}{\mathrm{E}(v|T > t, x)} \cdot \psi(t)\theta_0(x)$$

Clearly, because of the presence of unobserved heterogeneity (i.e., $\text{Var}(v) > 0$, which under regularity conditions implies that $\text{Var}(v|T > t, x) > 0$), the duration dependence in the observed (or "aggregate") hazard function $\theta(t|x)$ is more negative than otherwise. This is because in case of unobserved heterogeneity, the individuals with the highest values of v (and thus the highest hazards) on average leave the state quickest, so that the individuals who are still in this state at high durations tend to have lower values of v and thus lower hazards. This phenomenon has been called "weeding out" or "sorting". It occurs in duration models with unobserved heterogeneity in general, and so is not restricted to the MPH model. The model thus allows for two competing explanations for observed negative duration dependence. If one ignores the presence of unobserved heterogeneity (i.e. if one adopts a PH model whereas the data are generated by an MPH model with $\text{Var}(v) > 0$), then the estimated duration dependence will be too negative. This result has spurred the literature on the identification of duration models with unobserved heterogeneity.

Unobserved heterogeneity has a similar effect on the derivative of $\log \theta(t|x)$ with respect to x,

$$\frac{d \log \theta(t|x)}{dx} = \frac{\theta_0'(x)}{\theta_0(x)} - \frac{\text{Var}(v|T > t, x)}{\text{E}(v|T > t, x)} \cdot z(t)\theta_0'(x) \tag{20}$$

Note that in the case $\theta_0(x) = \exp(x'\beta)$, the first term on the right-hand side reduces to β, and $\theta_0'(x)$ in the second term reduces to $\theta_0(x)\beta$. Because of the presence of unobserved heterogeneity, the semi-elasticity of the observed hazard function $\theta(t|x)$ with respect to x is closer to zero than otherwise. This can be understood as follows. Within the group of individuals with a high value of $\theta_0(x)$, the weeding out induced by unobserved heterogeneity goes much faster than within the group of individuals with a low value of $\theta_0(x)$. This is a consequence of the multiplicative specification of $\theta(t|x, v)$: a high $\theta_0(x)$ and a high v reinforce each other in producing a very high hazard. As a result, at a given duration $t > 0$, the sample of survivors with high $\theta_0(x)$ has on average lower values of v than the sample of survivors with low $\theta_0(x)$. This causes the observed average difference between the hazards of the survivors of these groups to be smaller than the true average difference between the two groups. It is important to stress that this does not automatically imply that, if one ignores the presence of unobserved heterogeneity while estimating the model with Maximum Likelihood, that then the effect of x on the individual hazard is under-estimated. This is basically because β has one more element than x, and the ML estimates of β are jointly determined. We return to this in Subsection 5.6.

Note that if $\mathrm{E}(v) < \infty$ and $\psi(0) < \infty$ then for $t = 0$ the right-hand side of (16) reduces to $\psi(0)\theta_0(x)\mathrm{E}(v)$, and the function $\theta_0(x)$ is then identified from data on $\theta(0|x)$. This makes sense, as at $t = 0$ there is not yet any self-selection due to weeding out. Before we proceed with the identification of the full model (i.e., of the functions ψ, θ_0 and G), it is useful to introduce the function $h(s)$, defined as $-\mathcal{L}'(s)/\mathcal{L}(s)$ (see equation (18)). Equation (19) can now be rewritten as

$$\theta(t|x) = \psi(t) \cdot \theta_0(x) \cdot h(z(t)\theta_0(x)) \tag{21}$$

This equation will be useful in Subsection 5.3 and further.

5.2 Identification results

There is a substantial literature on the identification of the MPH model.[21] It is important to stress that no parametric functional form assumptions are made on the underlying functions θ_0, ψ and G, so the literature is concerned with *nonparametric* identification. In general it is assumed that the data provide the distribution function $F(t|x)$ for all t and x.

It is useful to define identifiability as a property of the mapping from the determinants ψ, θ_0 and G, given their domain, to the data (as summarized in $F(t|x)$ for all t and x). Consider a given set of assumptions on the three determinants (like the restriction that their function values must be nonnegative; below we examine various sets of assumptions). These characterize the domain of the mapping. The MPH specification then defines the unique mapping from the domain to the data. The model is identified if the mapping has an inverse, i.e. if for given data[22] there is a unique set of functions ψ, θ_0 and G in the domain that is able to generate these data.[23]

[21] Heckman (1991) provides an overview in which the MPH model is embedded in a more general class of models. Heckman and Taber (1994) list identification proofs for MPH models, non-MPH models, and more tightly specified MPH models without covariates.

[22] Of course, these data must be in the image of the mapping.

[23] In fact, for technical reasons, the identification literature typically focuses on the model determinant z instead of its derivative ψ.

Now let us consider the assumptions that are made on the determinants. These include the regularity Assumptions 1–4, and Assumption 5 on the independence of x and v. In addition, we list the following assumptions which will play a role in the remainder of the chapter,

Assumption 6 : Variation in observed explanatory variables. *The set $\mathcal{X}$ of possible values of x contains at least two values, and $\theta_0(x)$ is not constant on $\mathcal{X}$.*

Assumption 6b : Variation in observed explanatory variables. *There is an element x^a of the vector x with the property that the set $\mathcal{X}^a$ of its possible values contains a non-empty open interval. For given values of the other elements of x, the value of x^a varies over this interval. Moreover, $\theta_0(x)$ as a function of x^a is differentiable and not constant on this interval.*

Assumption 7 : Normalizations. *For some a priori chosen t_0 and x_0, there holds that $\int_0^{t_0} \psi(\tau)d\tau = 1$ and $\theta_0(x_0) = 1$.*

Assumption 8 : Tail of the unobserved heterogeneity distribution. $\mathrm{E}(v) < \infty$.

Assumption 8b : Tail of the unobserved heterogeneity distribution. *The random variable v is continuous, and the probability density function $g(v)$ of v has the property that*

$$\lim_{v \to \infty} \frac{g(v)}{v^{-1-\epsilon} S(v)} = 1 \tag{22}$$

where $\epsilon \in (0,1)$ is specified in advance, *and where $S(v)$ is a* slowly varying function,[24] *i.e. S has the property that, for every $v > 0$,*

$$\lim_{u \to \infty} S(uv)/S(u) = 1.$$

For Assumption 6, a single dummy variable x suffices, provided that it has an effect on the hazard function. In that case $\theta_0(x)$ takes on only two values on $\mathcal{X}$.

[24] See Feller (1971) for an exposition on such functions.

Note that we define θ_0 to be identified if its value is known for each $x \in \mathcal{X}$. In practice, one may start off with a parametric specification of $\theta_0(x)$ and require that all parameters can be recovered from the set of all pairs $(x, \theta_0(x))$ with $x \in \mathcal{X}$. In the case where $\theta_0(x)$ is (log-)linear in $x'\beta$, this implies that the elements of x should not be perfectly collinear.

Assumption 7 concerns an innocuous normalization of two of the three terms in the hazard $\theta(t|x, v)$. Assumptions 8 and 8b require more discussion. Basically, under Assumption 8, the right-hand tail of G is not allowed to be too fat because otherwise $\mathrm{E}(v) = \infty$. Now consider Assumption 8b. It is important to stress that the a priori choice of ϵ determines the assumed class of heterogeneity distributions. Basically, the smaller ϵ, the fatter the tails. However, for any $\epsilon \in (0, 1)$, all heterogeneity distributions have $\mathrm{E}(v) = \infty$ (see Ridder, 1990). This means that the right-hand tail of G is always fatter than under Assumption 8.

Elbers and Ridder (1982) were the first to prove the nonparametric identification of the MPH model, under Assumptions 1–8. Their identification proof is not constructive, i.e., the proof does not express the underlying functions θ_0, ψ and G directly in terms of observable quantities. Constructive identification proofs are attractive because they suggest a nonparametric estimation method. Melino and Sueyoshi (1990) provide a constructive proof for the case where Assumption 6 is tightened (to Assumption 6b, with the exception that $\theta_0(x)$ does not have to be differentiable). However, this proof is difficult to use as an inspiration for an attractive estimation strategy because it relies heavily on the observed duration density at $t = 0$, and x needs to be a continuous variable. Recently, Kortram et al. (1995) provide a constructive proof for the original case with only two possible values for $\theta_0(x)$. Lenstra and Van Rooij (1998) exploit this to construct a consistent nonparametric model estimator. They do not provide the asymptotic distribution of their estimator. Under somewhat stronger model assumptions than above, Horowitz (1999) constructs a nonparametric estimation method that does not follow an identification proof; rather, it exploits the similarity between the

MPH model and the transformation model (see Subsection 4.1).[25,26] He does provide the asymptotic distribution of his model estimator.

Heckman and Singer (1984b) also prove nonparametric identification of the MPH model. Their result turns out to be particularly interesting for the insights it generates into fundamental properties of the MPH model. Contrary to Elbers and Ridder (1982), they make Assumption 6b instead of the weaker Assumption 6, on the variation in x. More importantly, they make Assumption 8b instead Assumption 8 on the class of heterogeneity distributions. Assumption 8b rules out that v is degenerate. This means that the PH model as an underlying model is not included in the set of MPH models considered by Heckman and Singer (1984b). This is a disadvantage if the PH model is regarded to be an interesting special case. This result should *not* be taken to mean that the MPH models considered by Heckman and Singer (1984b) are not able to generate a PH specification for the *observed* hazard $\theta(t|x)$. Consider the set of MPH models generated by a particular choice of ϵ in (22), and assume that v has a Positive Stable distribution. This family of distributions is most easily characterized by its Laplace transform.

Family of Positive Stable distributions:

$$\mathcal{L}(s) = \exp(-s^{\alpha}), \qquad \text{with } \alpha \in (0,1).$$

Note that $\lim_{s\downarrow 0} \mathcal{L}'(s) = -\infty$, so $\mathrm{E}(v) = \infty$.[27] Using results in Ridder (1990) and Feller (1971) it can be shown that in fact we have to take α exactly equal

[25] In fact, Horowitz (1999) assumes that $\theta_0(x) = \exp(x'\beta)$, and he accordingly calls the estimator a semiparametric estimator. It should be stressed that this estimator and other nonparametric and semiparametric estimators for the MPH model rely heavily on the shape of the empirical survivor function for $t \downarrow 0$. For a number of reasons, it is notoriously difficult to assess this shape. For example, extremely short durations are often under-reported in real-life data.

[26] Horowitz (1999) also provides a useful list of existing semiparametric estimation methods where parametric functional forms are assumed for either ψ or G.

[27] The corresponding densities are bell-shaped (see Hougaard, 1986). Hougaard (1986) provides a justification of this family as a family of distributions for v in MPH-type models. Suppose that the individual duration can end for a number of different reasons $\{1,\ldots,n\}$, with cause-specific individual hazards that share the same baseline hazard and the same systematic hazard but not the same individual heterogeneity value v_j. The individual hazard, which is the sum of the cause-specific individual hazards, then equals $\sum \psi(t)\theta_0(x)v_j$, and this is an MPH specification with $v = \sum v_j$. Now suppose that the v_j are i.i.d. positive random variables, and suppose that $n \to \infty$. If the scaled mean of the v_j has a nondegenerate limiting distribution

to ϵ in order to obtain a G that satisfies (22). So, let v have a Positive Stable distribution with parameter ϵ. Then, by equation (19),

$$\theta(t|x) = \alpha\psi(t)[z(t)]^{\alpha-1}[\theta_0(x)]^{\alpha} \tag{23}$$

which is a PH specification, despite the fact that, according to the underlying model, there is unobserved heterogeneity. For example, if the underlying MPH model has a constant baseline hazard $\psi(t) = 1$ then the observed hazard has the (popular) Weibull PH specification with baseline hazard $\alpha t^{\alpha-1}$, with $0 < \alpha = \epsilon < 1$, which displays negative duration dependence.[28] Suppose that $\theta_0(x) = \exp(x'\beta)$. If the true model has a Positive Stable distribution of unobserved heterogeneity and if the researcher assumes instead that there is no unobserved heterogeneity and that $t|x$ has a PH specification (an assumption that is confirmed by the data!) then the parameter of interest β is estimated by $\beta\alpha$, so it is under-estimated in absolute value.

These results have very important implications. First, *the MPH model is non-parametrically unidentified if the assumption that $E(v) < \infty$ is dropped* (or, alternatively, if Assumption 8b is dropped). Moreover, the adoption of a model that is observationally equivalent to (but different from) the true model leads to biased inference on the parameters of interest (see also Robins and Greenland, 1989). This is bad news, as it is often difficult to make any justified assumption on the tail of the unobserved heterogeneity distribution. On the other hand, in the case where v represents an important economic variable, economic theory often provides a justification of $\mathrm{E}(v) < \infty$. In Subsection 5.5 we discuss some examples of this.

---

then it must be a Positive Stable distribution (Feller, 1971). In fact, for a wide range of distributions of the underlying random variable, the limiting distribution converges to a Positive Stable distribution. So, if v is an average of many different i.i.d. unobserved heterogeneity terms, then, in many cases, the distribution of v is approximated by a Positive Stable distribution. Note however that the underlying assumption that the different cause-specific hazards have the same baseline hazard and systematic hazard, while perhaps often reasonable in medical science, is often untenably strong in economics. Moreover, if v has a Positive Stable distribution and the parameter α is not fixed, then the MPH model is not identified (see below).

[28] If the underlying hazard has Weibull duration dependence $\psi(t) = (1/\alpha)t^{1/\alpha-1}$ and G is a Positive Stable distribution with parameter α then the observed hazard does not change with t, so $t|x$ has an exponential distribution.

Ridder (1990) addresses the fundamental identification problem in detail. He argues that for any MPH model with $\mathrm{E}(v) < \infty$ there are observationally equivalent models with $\mathrm{E}(v) = \infty$. In particular, for any MPH model with $\mathrm{E}(v) < \infty$ there is basically one observationally equivalent MPH model satisfying (22), for any $\epsilon \in (0, 1)$. So, Assumption 8 as well as Assumptions 8b for given ϵ can all be interpreted as different untestable normalizations that impose identifiability on a class of models that are unidentified.

Let us return to the case where v is degenerate (i.e., the PH model). Van den Berg (1992) proves that the full set of MPH models that is observationally equivalent to the PH model consists of models in which v is degenerate or has a Positive Stable distribution. In the latter case, as is clear from (23), the duration dependence of the baseline hazard and the absolute size of the effect of x are more positive than in the resulting PH model. For the general case, Ridder (1990) shows that some aspects of the MPH model are still identified if no assumptions on the tail of G are made. For example, the sign of the effect of x is identified.

As we shall see below, one solution to the fundamental identification problem is to rely on economic theory when choosing a functional form for G. Another solution is to use information on multiple spells for the same individuals.

5.3 Interaction between duration and explanatory variables in the observed hazard

In this subsection we examine properties of the observed hazard $\theta(t|x)$ if the underlying model has an MPH specification. These provide additional insights into the identification of the model. Throughout most of this subsection we assume that $\mathrm{E}(v) < \infty$, i.e. we adopt the MPH framework of Elbers and Ridder (1982). At times we generalize results by examining the wider class of models where $\mathrm{E}(v) \leq \infty$.

If there is no unobserved heterogeneity (so v is a constant), then the observed hazard $\theta(t|x)$ is multiplicative in t and x. Now suppose there is unobserved heterogeneity. If the observed hazard $\theta(t|x)$ would be multiplicative in t and x then the model would be observationally equivalent to a model without unobserved heterogeneity. Because of the nonparametric identifiability of the model, we know that the latter cannot be true. Therefore, the observed hazard cannot be multiplicative in t and x. As a result, we obtain the fundamental insight that *identification of G in MPH models comes from nonproportionality of the observed hazard* $\theta(t|x)$ (see Hougaard, 1991, Van den Berg, 1992, and Keiding, 1998). In terms of equation (21): if there is unobserved heterogeneity then the function $h(z(t)\theta_0(x))$ is not multiplicative in t and x, and the interaction between t and x identifies G. Yet another way to formulate this is by stating that if there is unobserved heterogeneity then $\log\theta(t|x)$ is not additive in t and x, so for some t and x

$$\frac{\partial^2 \log\theta(t|x)}{\partial t\partial x} \equiv \frac{\partial^2 \log h(z(t)\theta_0(x))}{\partial t\partial x} \neq 0 \tag{24}$$

provided that x varies continuously and the appropriate differentiability conditions are satisfied.

Now recall from the previous subsection that if the assumption that $\mathrm{E}(v) < \infty$ is dropped then a proportional specification for $\theta(t|x)$ can also be generated by MPH models with unobserved heterogeneity. Such models are characterized by the property that v has a Positive Stable distribution. All other distributions for v with $\mathrm{E}(v) = \infty$ generate $\theta(t|x)$ that is not multiplicative in t and x. Consequently, if Positive Stable distributions are ruled out for v then the result on the relation between unobserved heterogeneity and nonproportionality of the observed hazard can be extended to include infinite-mean distributions for v.

In fact, unobserved heterogeneity can not generate just any type of interaction between t and x in $\theta(t|x)$. Van den Berg (1992) shows that it is not possible that there are whole intervals of t and x on which there is no interaction.[29] (Whether the interaction is "large" is an empirical matter; as we shall see below, it is not difficult to construct examples in which there is virtually no interaction for a wide range of values of t.) Also, the following simple and appealing specification for $\theta(t|x)$ that allows for interaction cannot be generated with an MPH model,

$$\theta(t|x) = \psi(t)\theta_0(x)e^{-\alpha z(t)\theta_0(x)}$$

because the function $h(s) = \exp(-\alpha s)$ cannot be generated by the model.[30] In the next subsection we also derive restrictions on the sign of the interaction for

[29]This follows because any distribution G that gives a function h such that $h(z(t)\theta_0(x))$ is multiplicative in t and x on an interval must be a Positive Stable distribution.

[30]This can be seen as follows. If the model is an MPH model then $h(s)$ can be written as $-\mathcal{L}'(s)/\mathcal{L}(s)$, with $\mathcal{L}(s)$ being the Laplace transform of G. However, the function $\mathcal{L}(s)$ that follows from the candidate $h(s) = \exp(-\alpha s)$ is not completely monotone and hence cannot be a Laplace transform (see Feller, 1971).

different t. All of this evidence implies that the class of models for $\theta(t|x)$ that is generated by MPH models is smaller than the general class of interaction models for $\theta(t|x)$. In other words, the MPH model is overidentified. The fact that the function h must be such that it can be generated by a Laplace transform, the fact that $z(t)$ and $\theta_0(x)$ affect the value of h only by way of their product, and the fact that t enters the interaction term by way of the integral of the multiplicative term $\psi(t)$, all impose restrictions on $\theta(t|x)$ as a function of t and x.

At this stage it is instructive to examine the results in McCall (1996) on the identification of an extension of the MPH model with $\mathrm{E}(v) < \infty$ and $\theta_0(x) = \exp(x'\beta)$. Specifically, he allows the parameter β to vary with t. This is an empirically relevant extension (recall the discussion at the end of Section 4). Note however that the extension creates a second type of interaction between t and x in the observed hazard, so the question arises whether the data enable a distinction between them. McCall (1996) shows that the model is not identified if x can assume only two different possible values. However, if there is an explanatory variable that attains all possible values between $-\infty$ and ∞ then the model (i.e., ψ, G and $\beta(t)$) is identified, so then the two types of interaction can be distinguished empirically.

The inclusion of time-varying covariates (which is another empirically relevant extension of the MPH model) creates yet another type of interaction between t and x in the observed hazard. It is clear that in some cases a model with time-varying covariates is not identified (for example, if $\theta_0(x(t))$ is multiplicative in t). However, Honoré (1991) illustrates that in some cases time-varying covariates can also be helpful for identification. Suppose that x is time-invariant for part of the population; some of them have the value x_1 while others have x_2, with $\theta_0(x_2) \neq \theta_0(x_1)$. Suppose in addition that for the other part of the population the value of x changes discretely from x_1 to x_2 at duration $t^* > 0$, and assume that x satisfies the conditions for time-varying covariates laid out in Subsection 4.2. Then the model is identified without any assumption on the tail of G (so $\mathrm{E}(v)$ may be finite or infinite). See Heckman and Taber (1994) for a generalization of this result.

The results in McCall (1996), Honoré (1991) and Heckman and Taber (1994) illustrate the fact that the interaction generated by the presence of unobserved heterogeneity is rather specific. It is plausible that as more and more sources of interaction are included into the model, it becomes more and more difficult to achieve identification. In the limit, the assumption that the underlying hazard is multiplicative in t, x, and v is essential for identification. If this assumption is dropped then obviously any nonproportional specification can be generated without the need to allow for unobserved heterogeneity, and the model would be unidentified (see also Heckman, 1991). In particular, the specification (19) can also be generated as an individual hazard, which equals the observed hazard because of the absence of unobserved heterogeneity.

5.4 The sign of the interaction

In this subsection we examine the sign of the interaction between t and x in $\theta(t|x)$. This sign is a potentially interesting model characteristic, as its empirical counterpart may be readily observed from the data. Moreover, economic theory sometimes makes predictions of the sign of the interaction. For example, the ranking model of unemployment by Blanchard and Diamond (1994) predicts that the aggregate exit rate out of unemployment as a function of t decreases more in a "bad" steady state (i.e. a steady state where the exit rates are low anyway) than in a good steady state. If the steady state is represented by a dummy variable x then this means that the interaction between t and x is predicted to be always positive.

The discussion is facilitated by using $\theta_0(x)$ and x interchangeably. Obviously, this entails no loss of generality in the examination of the sign of the interaction, provided that it is kept in mind that x has a *positive* effect on $\theta(t|x, v)$. For convenience we take x to vary continuously, so that the sign of the interaction can be expressed as the sign of the cross-derivative of $\log h(z(t)x)$ with respect to t and x (see equation (24); recall that $\theta(t|x) = \psi(t) \cdot \theta_0(x) \cdot h(z(t)\theta_0(x))$).

The derivative of $\log h(z(t)x)$ with respect to x equals $h'(z(t)x)z(t)/h(z(t)x)$. The sign of the cross-derivative of $\log h(z(t)x)$ with respect to t and x then equals the sign of the derivative of $sh'(s)/h(s)$ evaluated at $s = z(t)x$. The function $h(s)$ is determined by the Laplace transform $\mathcal{L}(s)$ of G. Therefore, the sign of the interaction at a certain t and x is completely determined by G.[31] Given that $z(t)x$

[31] It follows from the results in Subsection 5.1 that $sh'(s)/h(s)$ at $s = z(t)x$ can be expressed in terms of the moments of $v|T > t, x$ (specifically, it depends on the first three moments).

takes on all values in $[0, \infty)$, knowledge of the sign of $sh'(s)/h(s)$ for all s is necessary in order to infer whether this sign is unambiguous for all t and x. To put this more bluntly, the full specification of the unobserved heterogeneity distribution determines the sign of the interaction between duration and explanatory variables in the observed hazard.

The first notable result concerns the sign of the interaction for small t. In general, the interaction is strictly negative on an interval $[0, \varepsilon)$.[32] This negative interaction means that if x is large then the observed duration dependence for small t is more negative than if x is small. This can be understood as follows. In the sub-population of individuals with a high value of x, the individuals who also have a high v will have a disproportionally high hazard. As a result, those individuals leave the state very quickly, and this has a strong negative duration-dependence effect on the observed hazard for the individuals with high x. Among the individuals with low x, this weeding out phenomenon occurs at a much lower speed, so their observed hazard decreases less strongly. It is important to stress that this intuitive explanation does not work for $t > 0$, because the distribution of v among survivors at $t > 0$ depends on x itself.

Lancaster (1979) shows that if G has a Gamma distribution,

Family of Gamma distributions:

$$g(v) = c^r / \Gamma(r) \cdot v^{r-1} \exp(-cv) \qquad \text{for all } v > 0, \text{ with } c, r > 0,$$

then the interaction is negative for all t and x, so the negative interaction sign for small t can be extended to all t. Unfortunately, this result cannot be generalized to include all possible G. To see this, consider discrete distributions for G with a finite number of mass points (or points of support), each of them positive and finite,

Family of discrete distributions with a finite number of mass points, each of them positive and finite:

$$\Pr(v = v_i) = p_i \quad \text{for all } i = 1, 2, \ldots, n,$$

$$\text{with } 0 < v_1 < v_2 < \ldots < v_n < \infty, 0 < p_1, p_2, \ldots, p_n < 1, \textstyle\sum_{i=1}^n p_i = 1, n < \infty$$

[32] For example, if $\text{Var}(v) < \infty$ and $\lim_{t \downarrow 0} \psi(t) \in (0, \infty]$ then $\partial^2 \log \theta(t|x)/\partial t \partial x < 0$ at $t = 0$. If $\text{E}(v^3) < \infty$ and $\lim_{t \downarrow 0} \psi(t) \in [0, \infty]$ then $\partial^2 \log \theta(t|x)/\partial t \partial x < 0$ on an interval next to $t = 0$.

(this is a popular specification in empirical work; see Subsection 5.5 below). We shall show that it is intuitively plausible that in this case, as $t \to \infty$, the derivative $\partial \log \theta(t|x)/\partial x$ goes to its value at $t = 0$ (so that this derivative varies with t in a non-monotone way, i.e. the cross-derivative does not have the same sign everywhere). When t increases, the group of survivors becomes increasingly more homogeneous, since the individuals with $v > v_1$ leave unemployment on average earlier than the individuals with $v = v_1$. In the limit, the group of survivors is homogeneous (all remaining individuals have $v = v_1$) so the value of $\partial \log \theta(t|x)/\partial x$ equals the value in a model without unobserved heterogeneity, which is $\theta_0'(x)/\theta_0(x)$ (see equation (20)). This in turn equals the value that is taken by $\partial \log \theta(t|x)/\partial x$ in general at $t = 0$ (see equation (20)), because at $t = 0$ the selection due to heterogeneity has not yet taken place.[33]

Example 7. Let v have a discrete distribution with two points of support with $\Pr(v = 1/5) = \Pr(v = 3/5) = 1/2$. Then the cross-derivative of $\log \theta(t|x)$ with respect to t and x equals zero if $z(t)\theta_0(x)$ is about 4.6 and it is positive if and only if $z(t)\theta_0(x)$ exceeds that number.

In this example, there is a positive value of $z(t)\theta_0(x)$ for which the observed hazard is multiplicative in t and x (i.e. the cross-derivative is zero) despite the presence of unobserved heterogeneity. However, the corresponding values of t and x have measure zero in the set of all possible values of t and x. Note that the above results implies that, if G is discrete with a finite number of points of support, the observed hazard $\theta(t|x)$ can be approximated by a PH specification if t is sufficiently large.

Incidentally, it is not difficult to construct examples where the weeding out of individuals with high v occurs very quickly after $t = 0$. If v has two points of support where one of them is extremely large, then the individuals with large v leave the state almost immediately. As a result, the *magnitude* of the interaction between x and t is virtually zero for almost all $t > 0$.

The family of discrete distributions is not the only family that generates a non-

[33] These results imply that, when comparing an individual with a relatively small x to one with a relatively large x, the proportionate difference between the observed hazards diminishes as time starts to run from $t = 0$ onward, but it ultimately returns to the level at $t = 0$.

monotone sign of the interaction. Other examples include uniform distributions with support $[c_1, c_2]$ with $0 < c_1 < c_2 < \infty$ as well as many other distributions with a positive lower bound of the support (see Abbring and Van den Berg, 1998, for details). In general, it seems difficult to derive conditions on G such that the interaction is always negative.[34] In the next subsection we return to this issue, when we examine the limiting distribution of $v|T > t, x$ as $t \to \infty$, for a wide class of distributions G.

Recall that in general for small t the interaction is negative. It turns out that, even if the interaction may be positive for larger t, the *cumulative* interaction remains negative. With this we mean that (under suitable regularity conditions),

$$\int_0^t \frac{\partial^2 \log \theta(\tau|x)}{\partial\tau\partial x}\, d\tau < 0$$

for all t and x. This can be seen by noting that this integral equals $\partial \log \theta(\tau|x)/\partial x$ at $\tau = t$ minus the same expression at $\tau = 0$, and, by equation (20), this is negative.

We end this subsection by noting a remarkable result on the effect of x on the observed hazard $\theta(t|x)$ in MPH models.[35] One may be tempted to think that this effect is always positive if x has a positive effect on the underlying hazard $\theta(t|x, v)$. However, this is not a general property of the model. Intuitively, if a fraction of individuals has a very high value of v then, in the sub-population of individuals with high x, the high-v individuals leave the state extremely quickly. The drop in the mean value of v among the survivors with high x is then so large that their hazard may on average fall below the value of those with lower x values. In such a case, the negative effect of the drop in v on $\theta(t|x)$ is not offset by the positive effect of the large x. In terms of equation (20), the second term on the right-hand side dominates the first one.

Example 8. Consider again the discrete distribution for v with $\Pr(v = 1/5) = \Pr(v = 3/5) = 1/2$ (see Example 7). Then $\partial\theta(t|x)/\partial x$ is always positive. However,

[34] Negative interaction is equivalent to the statement that $-\mathcal{L}'(e^y)/\mathcal{L}(e^y)$ is log-concave on $y \in (-\infty, \infty)$, but this does not seem to correspond to a well-known class of distributions for G.

[35] Even though this result is not concerned with the sign of the interaction, its interpretation fits in with the latter subject.

if the highest mass point is at 5/2 instead of 3/5 the this derivative is negative for values of t and x such that $z(t)\theta_0(x)$ is in an interval around 1.

In sum, the observed hazard of a high-x individual can be smaller than that of a low-x individual. This means that it is not possible to deduce the sign of the effect of x on the underlying individual hazard from the observed relation between x and the observed hazard at a certain duration t. It should however be stressed that this remarkable effect can only occur for some local duration intervals. Specifically, the observed survivor function $\overline{F}(t|x)$ and the observed mean duration $\mathrm{E}(t|x)$ are always decreasing in x (iff $\theta_0(x)$ increases in x). This can be seen from the relations

$$\overline{F}(t|x) = \mathrm{E}_v(\overline{F}(t|x,v)) = \mathcal{L}(z(t)\theta_0(x))$$

$$\mathrm{E}(t|x) = \mathrm{E}_v\mathrm{E}(t|x,v) = \int_0^\infty \mathcal{L}(z(t)\theta_0(x))dt$$

where E_v denotes the expectation with respect to G (note that $\mathcal{L}$ decreases in its argument; see equation (18)).

5.5 Specification of the unobserved heterogeneity distribution

Studies in which parameterized MPH models are estimated have wrestled with the choice of a functional form for G (see e.g. Heckman and Singer, 1984a). This choice is thought to be harder to justify than the choice for a functional form for the baseline hazard ψ, as economic theory often suggests a shape for the latter. In this subsection we examine parametric families of distributions that can be given supporting arguments as a choice for G. We start with families that can be supported by limit arguments. Next we show that economic theory sometimes actually does make informative predictions on important aspects of the shape of G. This typically concerns cases where a key source of individual heterogeneity is observed by labor market participants but not by the researcher.

5.5.1 Discrete distributions

Suppose that the baseline hazard and the systematic hazard have parametric functional forms with a finite number of parameters, but that the only assumption on G is that it has a finite mean (or satisfies (22)). For this case, Heckman and Singer (1984c) show that the Maximum Likelihood estimator of G is a discrete distribution, provided that some regularity conditions are met.[36] For a given sample, the parameters of this discrete distribution (the number of points of support, their location, and their associated probabilities) are chosen such as to maximize the likelihood function. The result by Heckman and Singer (1984c) illustrates the flexibility of discrete distributions as heterogeneity distributions. Intuitively, if the number of points of support increases, then any true underlying distribution G can be approximated well. In practice, it is often difficult to find more than a few different mass points. Usually, if more than two or three points of support are taken then the estimates of some of them coincide. Standard practice in case of discrete G is to estimate the model with a number of mass points that is either predetermined or equal to the maximum number that could be detected, and to report standard errors conditional on this choice. It is important to stress that such approaches are not "nonparametric" in the true sense of the word, and that the standard errors do not reflect uncertainty with respect to the actual number of mass points.

The fact that it is often difficult to find more than a few mass points may reflect a lack of informativeness on G in the data. Recall that the data do not provide observations on drawings from G, but that G enters the likelihood function as a mixing distribution. The information on G comes from the observed interaction between t and x in the data, and it may be that a mixing distribution with a few mass points is often able to capture most of this. The simulations in Heckman and Singer (1984c) strongly confirm this. They find that the parameters of ψ and θ_0 as well as the shape of the distribution of $t|x$ are well estimated if G is assumed to be discrete with an unknown number of mass points, even if the true G is continuous. The estimated number of mass points is typically small.

For G discrete with a finite number of points of support, each of them positive and finite, we restate the following model properties. First, $\mathrm{E}(v) < \infty$. Secondly, the interaction between t and x in $\theta(t|x)$ is not monotone; it is negative for small t and positive for very large t. Thirdly, the effect of x on $\theta(t|x)$ is not always monotone even if the effect on $\theta(t|x, v)$ is.

[36] See Trussell and Richards (1985), Lancaster (1990) and Baker and Melino (2000) for additional insights into this estimator and for alternative computational strategies.

5.5.2 Gamma distributions

In applications, the family of Gamma distributions has perhaps been the most popular choice for G. This stems from the resulting analytic tractability: all relevant properties of the distribution of $t|x$ can be expressed in closed-form solutions. In their recent working paper, Abbring and Van den Berg (1998) are the first to provide a less ad-hoc justification for the choice of the family of Gamma distributions for G. Suppose that zero is the lower bound of the support of the true (unknown) G, with v being a continuous random variable (we do not make assumptions on the upper bound of the support of G). Then, under mild regularity conditions, the unobserved heterogeneity distribution among the survivors at duration t converges to a Gamma distribution if $t \to \infty$. In fact, we have to scale the distribution of v among survivors because the unscaled distribution converges to zero (note that the Gamma family is invariant to scaling). This result implies that, in many cases, the heterogeneity distribution among survivors at high durations can be approximated well by a Gamma distribution, and this provides a motivation to adopt the Gamma family for $G(v)$ itself.

For $G(v)$ equal to a Gamma distribution, we restate the following model properties. First, $\mathrm{E}(v) < \infty$. Secondly, the interaction between t and x in $\theta(t|x)$ is monotone and negative for all t. Thirdly, the effect of x on $\theta(t|x)$ is always monotone if the effect on $\theta(t|x,v)$ is monotone.

The limit result in Abbring and Van den Berg (1998) does not hold if the true $G(v)$ is a discrete distribution with a finite number of points of support.[37]

5.5.3 Suggestions from economic theory

Now let us turn to (aspects of) shapes of $G(v)$ that can be justified by economic theory. First, as a general remark, it should be noted that economic theory often predicts that the exit rate out of a state is bounded from above. Consider the search theories of Section 3. In general, the exit rate out of unemployment can be written as $\lambda \overline{F}(\phi)$. The second term in this expression is a probability which

[37] Recall that in such a case the sign of the interaction is positive for large t, whereas in the case of a Gamma distribution it is negative for large t. The latter suggests that, if in practice a choice must be made between a discrete G or a Gamma G, it is useful to examine the sign of the interaction between t and x in the data on $\theta(t|x)$ for large t (see Hougaard, 1991, for an example).

necessary lies between zero and one. If the first term is infinite then there are no frictions in the first place, and the models reduce to standard labor market models with zero unemployment durations. According to this line of reasoning, $\theta(t|x, v)$ should be bounded from above, which implies that the support of G is bounded from above (which in turn implies that $\mathrm{E}(v) < \infty$).[38]

Suggestions from equilibrium search models

Suppose worker behavior is described by the search models of Section 3. In the literature, these models have been extended to include employer behavior. For surveys of the theoretical and empirical analysis of such "equilibrium search models", see Ridder and Van den Berg (1997), Mortensen and Pissarides (1999), and Van den Berg (1999). To fix thoughts, consider the equilibrium search model of Bontemps, Robin and Van den Berg (1999) where unemployed and employed workers search, and different workers have different values of leisure b. If the job offer arrival rates are the same in employment and unemployment, then the reservation wage of an unemployed worker with value of leisure b is simply equal to b. Now suppose that b has a continuous distribution $H(b)$ in the population. An employer sets his wage w such as to maximize his steady-state profits. We assume that the number of firms is fixed, or, alternatively, that an entry fee has to be paid. It is not optimal for any firm to offer a wage equal to the lower bound $\underline{b}$ of the distribution $H(b)$, because then its steady-state labor force and profit rate are zero. The lowest wage $\underline{w}$ in the market is strictly larger than $\underline{b}$. As a result, there is a positive fraction of individuals who accept any wage offer (i.e., who have $b < \underline{w}$).

In this model, the individual exit rate out of unemployment equals $\lambda\overline{F}(b)$. Now suppose that the researcher wants to estimate a reduced-form model of unemployment durations. The individual value of leisure b is unobserved, so it is reasonable to take the unobserved heterogeneity term v to represent the acceptance probability $\overline{F}(b)$ (provided that there is no additional source of unobserved heterogeneity). As a result, the distribution $G(v)$ has support in $[0, 1]$. But there

[38] One may argue that λ is affected by an optimally chosen search intensity, and that the distribution of structural determinants in the population is such that the resulting distribution of λ does not have an upper bound. However, in search and matching models, λ is at least partially determined by the meeting technology of the labor market; this technology is a market characteristic that cannot be fully dominated by individual behavior.

is a positive fraction of workers with $\overline{F}(b) = 1$, so G has a mass point at the upper bound of its support (i.e., at $v = 1$). If the highest wage in the market $\overline{w}$ is smaller than the highest level of b then G also has a mass point at zero. In that case G is a defective distribution; a positive fraction of individuals is unemployed forever. In practice it may not be difficult to sort out the latter individuals from the data (i.e. to observe whether $b > \overline{w}$), because it does not make sense for these individuals to search for a job, so they may classify themselves as being nonparticipants.

It is not difficult to see that this result extends to more general equilibrium search models. Often, employer behavior is such that a positive fraction of unemployed workers accepts any wage offer and consequently has the maximum hazard level for the transition into employment.

Suggestions from on-the-job search models

Consider the stationary on-the-job search model of Subsection 3.2. Published statistics on nationwide job mobility contain information on the marginal job duration distribution, i.e. on the distribution of job durations unconditional on the wage in the job. The wage then represents unobserved heterogeneity in the job duration data.

The distribution of t given the wage w on the job is exponential with density

$$f(t|w) = (\delta + \lambda_1 \overline{F}(w))e^{-(\delta+\lambda_1\overline{F}(w))t} \tag{25}$$

Consider the job durations t of a cohort of workers who have just left unemployment for a job (this constitutes the inflow into employment at a given point of time). If all unemployed workers accept any wage that is offered to them then, in this cohort, the wage w is distributed according to $F(w)$. To obtain the marginal job duration distribution for this cohort, we have to integrate (25) with respect to $dF(w)$. This gives

$$f(t) = \frac{1}{\lambda_1}\int_{\delta}^{\delta+\lambda_1} ze^{-zt}dz$$

which is a "mixture of exponentials" *i.e.*, a mixture of distributions with constant hazards, with a uniform mixture distribution for the hazards with support on the interval $(\delta, \delta + \lambda_1)$.[39] This is not surprising. The conditional hazard of $t|w$ is constant over the job duration. It is then mixed with respect to a determinant (w) of the conditional hazard. Workers are merely concerned with the ordering of the current wage and the wage offer, and not with the shape of the underlying wage offer distribution itself. Their location on the job ladder therefore determines their hazard. Note that, as a result, the marginal job duration distribution does not depend on F.

In terms of an MPH model, $\theta(t|x)$ can be thought of as being generated by $\theta(t|x, v) = v$, where v has a uniform distribution on $(\delta, \delta + \lambda_1)$.[40] This result for a cohort of newly employed workers can be generalized to other (more relevant) sampling schemes. Ridder and Van den Berg (1998) apply this approach to study job mobility with aggregate data.

The argument above also applies to other settings where only the rank of the individual's heterogeneity value affects the individual's hazard rate, and where these values and their ranks are unobservable. Moscarini (1998) examines a job search model for the unemployed where individuals are ranked by employers on the value of some time-invariant characteristic. The rate at which an individual obtains a job depends on the fraction of the unemployed that has worse characteristics. For a specific matching technology, this results in an unemployment duration distribution that is again a mixture of exponential distributions with a uniform mixture distribution.

5.6 Effects of misspecification of functional forms

Generally, in applications, ψ and/or $G(v)$ are assumed to have a parametric functional form (see Lancaster, 1990, for a catalogue of popular functional forms). We finish this section on properties of the MPH model by summarizing some results on the effects of misspecification of these functional forms on the probability limits of the Maximum Likelihood (ML) estimates. Throughout the subsection (and

[39] This can be further simplified to

$$f(t) = \frac{e^{-\delta t}}{\lambda_1 t^2}\left[1 + \delta t - (1 + (\delta + \lambda_1)t)e^{-\lambda_1 t}\right].$$

[40] Note that, if δ or λ_1 depend on t or x, then this is not an MPH model anymore.

in line with this literature) we assume that

$$\theta_0(x) = \exp(\beta_0 + x'\beta_1)$$

and that all moments of v exist. The model is normalized by taking $\mathrm{E}(v) = 1$. The only type of censoring that is considered concerns independent right-censoring at a fixed duration.

A natural starting point concerns the misspecification due to omission of unobserved heterogeneity from the model, if it is present in the data-generating process. Recall that in Subsection 5.1 we argued that the estimated duration dependence will be too negative, and the effect of x may be inconsistently estimated as well. Gail, Wieand and Piantadosi (1984) provide the following result. If the baseline hazard $\psi(t)$ is known a priori, if one erroneously ignores unobserved heterogeneity in the model specification, *and* if there is no censoring, then β_1 is consistently estimated with ML. In fact, it is not difficult to show that

$$\mathrm{plim}\widehat{\beta}_0 = \beta_0 - \mathrm{E}(\frac{1}{v}) < \beta_0, \qquad \mathrm{plim}\widehat{\beta}_1 = \beta_1$$

where $\mathrm{plim}\widehat{\beta}_i$ denotes the probability limit of the ML estimator of β_i (i.e., the value to which the estimate converges in probability as the sample size increases). Note that $\mathrm{E}(1/v) > 1/\mathrm{E}(v) = 1$ if and only if $\mathrm{Var}(v) > 0$, i.e. if there is unobserved heterogeneity.[41,42]

Unfortunately, these welcome results do not generalize in any way to more realistic settings. Ridder (1987) shows that censoring in the data makes $\widehat{\beta}_1$ inconsistent (unless the specified G equals the true G or $\beta_1 = 0$). The asymptotic bias is towards zero if the specified model assumes absence of unobserved heterogeneity.

[41] See also Lancaster (1983). Ridder (1987) generalizes this result by proving the following: if the baseline hazard is known in advance, the assumed G is fully specified without unknown parameters, the assumed G is not equal to the true G, and there is no censoring, then β_1 is consistently estimated.

[42] This is not in conflict with the result in Subsection 5.1 that $d\log\theta(t|x)/dx = \beta_1(1-a)$ for some $a > 0$. Somewhat loosely one may say that $\widehat{\beta}_0$ ensures that the average level of the specified $\log\theta(t|x)$ agrees to the average level in the data, and that the effect of x in the data is best captured by $\widehat{\beta}_1 = \beta_1$. Note that in this specific model, $\mathrm{E}(\log z(t)|x, v)$ is additive in v and x. In particular, $\mathrm{E}(\log z(t)|x, v) = -\beta_0 - x'\beta_1 - \log v + c$, with $c \approx -0.58$ being the mean of an EV1 random variable, and with the function $z(.)$ completely known. So by analogy to the regression model, dispersion in v does not affect the estimate of β_1.

Lancaster (1985b) shows that if the baseline hazard is known to have a Weibull specification with an unknown parameter, one ignores unobserved heterogeneity, and there is no censoring, then the estimates of both the Weibull parameter and β_1 are asymptotically biased towards zero. In fact, they are all biased in the same proportion. Basically, in this case, ML gets the regression function for $\log t$ right, but we are after the original parameters of the individual hazard function instead of the elasticities of the mean log duration. Ridder (1987) also shows that misspecification of the shape of the baseline hazard results in inconsistency of $\widehat{\beta}_1$.

The results above are all analytically derived. For more general model settings, the effects of misspecification have been analyzed by way of extensive Monte Carlo simulations. Ridder (1987) allows for censoring in the Lancaster (1985b) model, and he allows for misspecified G in the assumed model. It turns out that censoring exacerbates the asymptotic bias in $\widehat{\beta}_1$ due to misspecification of G, and the results become sensitive to the assumed specification of G. Moreover, it turns out that the estimates display a large small-sample bias even if the model specification is correct. This bias disappears very slowly when the sample size increases. Such small-sample biases are absent for the PH model without unobserved heterogeneity; see Andersen, Bentzon and Klein (1996).

Ridder (1987) also examines the performance of ML estimation of an assumed model with a Weibull baseline hazard and a Gamma distribution for v, if both are misspecified. The simulations reinforce the negative results above. Ridder (1987) conjectures that if the baseline hazard is flexibly specified with a sufficient number of unknown parameters, and if censoring is virtually absent, then it does not matter which family of distributions is assumed for G in order to obtain a reliable estimate of β_1. However, the simulation results in Baker and Melino (2000) go against this.[43] Most of the biases due to the above problems can be substantial, depending on the situation at hand. For the Partial Likelihood estimation method, similar results have been derived (see e.g. Bretagnolle and Huber-Carol, 1988).

By now there are also many studies of real-life single-spell data in which it is reported that the estimates of (the parameters of) β_1, ψ and G are sensitive to

[43] It should be noted, though, that Baker and Melino (2000) do not examine an MPH model but a discrete-time model where the individual per-period exit probability is a logistic function of $\psi(t)\theta_0(x)v$. Whether these models behave similarly is an issue for further research.

changes in the assumed family of distributions for G or the assumed set of x or the assumed functional form of ψ, even though sometimes the over-all fit of the model does not change with this in any substantial way (see e.g. Heckman and Singer, 1984a, Trussell and Richards, 1985, Hougaard, Myglegaard and Borch-Johnsen, 1994). Keiding, Andersen and Klein (1997) provide a survey of studies with biostatistical data.

The recent literature on semiparametric and nonparametric estimation of the MPH model provides some interesting additional insights on this. First of all, Hahn (1994) examines models with Weibull duration dependence, and he assumes that v is a continuous random variable with a finite mean. He shows that with single-spell data, the information matrix is singular, and that there is no $\sqrt{n}-$consistent estimator for β_i and the Weibull parameter.[44] Thus, in a certain sense, there is less information on the model parameters than what is typically available in econometric analyses. Secondly, Heckman and Taber (1994) and Kortram et al. (1995) show that the mapping from the data-generating process to the data is not continuous, so that two distinct MPH models can generate very similar data.[45] Thirdly, the nonparametric (or semiparametric) estimator developed by Horowitz (1999) has convergence rates that are smaller than $\sqrt{n}$. In particular, under certain assumptions (including absolute continuity of an element of x, differentiability of $\psi(t)$ and the density of v, and $\mathrm{E}(v^2) < \infty$), the convergence rates of β_i and ψ can be at most almost equal to $n^{-2/5}$, which is obviously slower than $n^{-1/2}$. For the heterogeneity distribution and density G and g, the rate of convergence is $(\log n)^{-2}$, which is *very* slow.

Together, these results lead to the following conclusion. In the absence of strong prior information on the determinants of the MPH model, single-spell data do not enable a robust assessment of the relative importance of these determinants as explanations of random variation in the observed durations (even

[44] See Klaassen and Lenstra (1998) for a generalization of this result.

[45] As an example, consider the simplest MPH model, with $\theta_0(x) = \exp(x)$ where x is a single dummy variable, and with absence of duration dependence and unobserved heterogeneity. The distribution of $t|x$ is virtually the same as the distribution generated by an MPH model with $\theta_0(x) = \exp(2x)$, duration dependence proportional to $2t$, and v distributed as a Positive Stable distribution with parameter $1/2$ with the upper tail replaced by a finite mass point (see Kortram et al., 1995, for details; note the similarity to the example in the discussion in Subsection 5.2; also note that here $\mathrm{E}(v) < \infty$).

if the unobserved heterogeneity mean is known to be finite). Minor changes in the assumed parametric specification, leading to a similar over-all fit, may produce very different parameter estimates. This implies that estimation results from single-spell data are sensitive to misspecification of the functional forms associated with these determinants. Therefore, interpretations based on such results are often unstable and should be performed with extreme caution.

In biostatistics, this state of affairs has led to a renewed interest in Accelerated Failure Time models for the analysis of single-spell duration data (see Hougaard, Myglegaard and Borch-Johnsen, 1994, and Keiding, Andersen and Klein, 1997, for a survey). Note that such models allow for robust inference on the effect of x on the mean of $\log t$.[46] In a way, the choice for the AFT model means that all hope is given up on the attempt to (i) disentangle genuine duration dependence from the effect of unobserved heterogeneity, and (ii) quantify the effect of covariates on the *individual* hazard as opposed to the observed hazard, with single-spell data. From an economic-theoretic point of view, however, the AFT approach is unsatisfactory, because, as we have seen in Sections 2 and 4, the parameters of the individual hazard are the parameters of interest. It may therefore be better to exploit predictions from the underlying economic theory when specifying the duration model, and/or look for data with multiple spells.[47]

If one is only interested in the *sign* or *significance* of a covariate effect on the individual durations then the AFT approach may be useful. Recall from Subsection 5.4 that in MPH models the sign of the effect of x on the mean duration is always the same as the sign of the effect on the individual hazard, regardless of the specification of ψ or G. Regression of $\log t$ on x therefore provides robust evidence on this sign (see Solomon, 1984, 1986, for proofs; Li, Klein and Moeschberger, 1993, provide supporting Monte Carlo evidence on the performance of test statistics for the significance of the effect of x). Such an approach may be useful if one is interested in whether participation in a treatment program (to be repres-

[46] Indeed, Horowitz (1996) shows that the β parameters in the transformation model (9) can be consistently estimated with an estimator with convergence rate equal to $n^{-1/2}$. Recall that the AFT model is a special case of the transformation model.

[47] Another approach would be to estimate the model nonparametrically using methods described in Subsection 5.2. It is still too early to assess whether this approach is fruitful. Yet another approach is to use population data (if available). See Van den Berg and Van Ours (1996) for an example of this based on a discrete-time model.

ented by x) has any effect. However, in economics, data on treatment effects are usually non-experimental and treatment assignment is selective, so then x is not exogenous (see Subsection 9.2).

6 The MPH model with multi-spell data

6.1 Multi-spell data

This section deals with identification of the MPH model if the data provide durations of multiple spells in a given state by a given individual, i.e. if the data are *multi-spell* data. Here, an individual has a given value of v, and his spell durations are independent drawings from the univariate duration distribution $F(t|x,v)$, where, of course, v is unobserved, so that the durations given just x are not independent. We mostly focus on an "ideal" case in which the data consist of a random sample of individuals and provide two uncensored durations for each individual in the sample. Actually, the use of the term "individual" is not very appropriate here, as the setup includes cases in which physically different individuals are assumed to share the same value of v and we observe one or more durations for each of these individuals. It is convenient to refer to such a group of individuals as a *stratum*. It depends on the context whether one may assume that v, ψ, and θ_0 are identical across durations for the same individual or stratum. In subsequent sections we examine more general models, in which ψ and θ_0 may vary across spells, the values of v in different spells may be stochastically related, and other dependencies between the durations are allowed. It is useful to think of the present section as being concerned with a model for a single type of duration, where we have multiple spells of this type of duration for each "individual", whereas the subsequent sections are concerned with models for different types of durations with single or multiple spells of each type for each "individual".

The empirical analysis of MPH models with multi-spell duration data is widespread. For example, Newman and McCullogh (1984) use such data to estimate reduced-form models for birth intervals, while Ham and Rea (1987) and Coleman (1990) use such data to estimate reduced-form unemployment duration models.[48] Lillard (1993) and Lillard and Panis (1996) estimate marriage duration models

[48] Ham and Rea (1987) use a discrete-time model.

with multi-spell data. In these applications, the multiple spells with a given value of v are associated with a single physical individual. There are also many applications in which multiple spells with a given v are associated with different physical individuals (see e.g. Kalbfleisch and Prentice, 1980). The heterogeneity term is then assumed to be identical across individuals within some group or stratum. Typically, different individuals within a stratum are allowed to have different values of x. As we shall see below, this may actually be very useful for inference.[49] Recent applications include Guo and Rodríguez (1992), Wang, Klein and Moeschberger (1995), Sastry (1997), Ridder and Tunalı (1999), and Lindeboom and Kerkhofs (2000). Arroyo and Zhang (1997) survey applications in the analysis of fertility. In studies on lifetime durations of identical twins, the unobserved heterogeneity terms are often assumed to capture unobserved genetic determinants, so then v is identical within twin pairs (see e.g. Hougaard, Harvald and Holm, 1992a).

To proceed, note that the individual hazard function $\theta(t|x,v)$ is the same for both durations associated with the "individual". The value of x may differ between the corresponding spells. If necessary we denote the values by x_1 and x_2, respectively. Conditional on x and v, the two durations t_1 and t_2 are independent. Conditional on x, the variables t_1 and t_2 are independent if there is no unobserved heterogeneity, i.e. if v is not dispersed.

If $\theta_0(x) = \exp(x'\beta)$ then

$$\begin{aligned} \log \int_0^{t_1} \psi(u)du &= -x_1'\beta - \log v + \varepsilon_1 \\ \log \int_0^{t_2} \psi(u)du &= -x_2'\beta - \log v + \varepsilon_2 \end{aligned} \tag{26}$$

where ε_1 and ε_2 are i.i.d. EV1 distributed. Equations (26) suggest a similarity to standard panel data models with fixed effects. We return to this below.

The joint density $f(t_1,t_2|x)$ of t_1 and t_2 given x can be expressed as

$$f(t_1,t_2|x) = \int_0^\infty \int_0^\infty f(t_1|x_1,v)f(t_2|x_2,v) \quad dG(v) \tag{27}$$

[49] Indeed, with stratified partial likelihood inference, estimation of the systematic hazard θ_0 is *driven* by the variation in x (see Subsection 6.2).

in which G denotes the joint distribution of v across "individuals" in the population. The density $f(t_i|x_i,v)$ can of course be expressed in terms of the determinants of θ (see Section 2). The joint survivor function of t_1 and t_2 given x can then be expressed as

$$\overline{F}(t_1,t_2|x) = \int_0^\infty e^{-[z(t_1)\theta_0(x_1)+z(t_2)\theta_0(x_2)]v} dG(v)$$

In many applications, the individual likelihood contribution is based on the density (27). In terms of panel data analysis, this means that the values of v are treated as "random effects" when estimating the model with Maximum Likelihood.[50] An alternative empirical approach treats v as individual-specific parameters or "incidental" parameters. The likelihood function is then written for given unknown values of these (and the other) parameters.[51]

6.2 Identification results

One may distinguish between two approaches in the literature on identification of the MPH model with multi-spell data. The first approach below is concerned with the full identification of the model and relies on results that were discussed in Section 5. The second approach is concerned with the identification of the systematic hazard θ_0 and follows from properties of a particular estimation method.

We start with the first approach. Honoré (1993) shows that the MPH model with multi-spell data is identified under much weaker assumptions than in Section 5. In fact, we do not need to assume that there are observed explanatory variables x at all. In other words, the analysis is conditional on a given value of x, and we may allow for full interaction of the actual value of x with the model determinants: ψ may depend on x in an unspecified way, and v and x may be dependent in the population. Note that here x does not vary across spells for a given individual. We may write

$$\theta(t|x,v) = \psi(t|x)\cdot v, \qquad v|x \sim G(v|x)$$

This includes of course as a special case that $\psi(t|x)$ can be written as $\psi(t)\theta_0(x)$.

[50] Here, as in the model with single spells, standard maximization of the likelihood may be computationally unfeasible for particular parametric specifications for G and ψ. In such cases, use of the EM algorithm may be preferable (see Lancaster, 1990, for details).

[51] See Lancaster (2000a) for a general overview of incidental parameters in econometrics.

This model is identified given regularity assumptions corresponding to Assumptions 2–4, and given a normalization of the integrated baseline hazard (analogical to Assumption 7). Thus, if two observations are available for each v, then the identification of the model does not require an untestable assumption on the tail of the unobserved heterogeneity distribution G anymore, and, perhaps even more importantly, v and x are allowed to be dependent. The identification of this distribution does not come anymore from the interaction between the duration and the observable explanatory variables in the observed hazard. The identification does however need proportionality of the duration effect and the unobserved heterogeneity term in the individual hazard. It should be noted that this model is nevertheless overidentified; see Subsection 8.2.2.

Example 9. Let $\psi = 1$ (so there is no duration dependence) and $x_1 = x_2 (= x)$, and suppose that v has a Positive Stable distribution (see Subsection 5.2). Such distributions have infinite means. As we have seen, the resulting MPH model for single spells is observationally equivalent to a PH model without unobserved heterogeneity and a Weibull baseline hazard. However, it is easy to see that the joint survivor function of t_1 and t_2 equals

$$\overline{F}(t_1, t_2|x) = \exp\left(-[\theta_0(x)]^{\alpha}(t_1 + t_2)^{\alpha}\right)$$

(with $0 < \alpha < 1$), whereas if there is no unobserved heterogeneity and the baseline hazard has a Weibull specification ($\psi(t) = \alpha t^{\alpha-1}$) then

$$\overline{F}(t_1, t_2|x) = \exp\left(-[\theta_0(x)]^{\alpha}(t_1^{\alpha} + t_2^{\alpha})\right)$$

so the two models are observationally distinct, even if $\theta_0 = 1$.

Now let us turn to the second approach to identification, which focuses on the effect of observed explanatory variables on the individual hazard function. The systematic hazard θ_0 is identified under very weak conditions if the data contain multiple spells with the same value of v. This has been known for some time, for the reason that a nonparametric estimation method exists for θ_0 in this setup (see Kalbfleisch and Prentice, 1980, and Chamberlain, 1985). In fact, this estimation method is applicable to a model setup that is more general than the MPH model. To proceed, it is useful to distinguish between observed explanatory variables x^*

which do not vary within strata, and observed explanatory variables x which do vary within strata. We assume for expositional reasons that the hazard function is multiplicative in a part depending on x^* and a part depending on x. In particular,

$$\theta(t|x^*, x, v) = \psi(t|x^*, v) \cdot \theta_0(x), \qquad v|x^*, x \sim G(v|x^*, x) \tag{28}$$

This specification allows for full interaction of the values of v and x^* with the elapsed duration t in the hazard function. This implies that we allow the baseline hazard to differ across strata (i.e., across groups of spells with the same v). Moreover, v, x^* and x may be dependent. The basic idea of the estimation method is that a Cox partial likelihood can be constructed *within* strata. For a given stratum, the partial likelihood depends only on θ_0, and not on G or ψ or the values of v or x^*. These likelihoods can be combined to construct an over-all partial likelihood which can be used to estimate θ_0 (see the above references for details).

Clearly, the effects of the explanatory variables x^* cannot be estimated from this. In other words, to be able to estimate the effect of an observed explanatory variable with this approach, it is essential that the values of the variable sometimes differ across spells within a stratum. In case of two spells per stratum, this amounts to $x_1 \neq x_2$. To see this, note that within such a stratum,

$$\Pr(t_1 > t_2|x_1, x_2, v) = \frac{\theta_0(x_2)}{\theta_0(x_1) + \theta_0(x_2)}$$

which is only informative on θ_0 if $x_1 \neq x_2$.

The within-stratum baseline hazard ψ as a function of t can subsequently be estimated nonparametrically. Yamaguchi (1986) surveys these methods. Kalbfleisch and Prentice (1980) and Ridder and Tunalı (1999) contain useful expositions on the inclusion of time-varying covariates.

What does this "stratified partial likelihood" estimation approach imply for the identification of θ_0 in the MPH model with multi-spell data? This function is identified up to a multiplicative constant if θ_0, ψ, and G in equation (28) satisfy regularity assumptions corresponding to Assumptions 1–4, and if x varies between spells within strata. Again, we do not need independence of observed and unobserved explanatory variables, and we do not need an assumption on the tail of the distribution of the unobservables. Note that the identification result is valid under a specification of the hazard function that is much more general than the MPH specification.

The approach of the previous paragraphs is particularly appealing if the individual v are regarded as incidental parameters. With full ML, such parameters can in general not be estimated consistently if asymptotically the number of strata goes to infinity with a fixed number of spells per stratum (Lancaster, 2000a). In the above approach, however, these parameters cancel out of the partial likelihood. Somewhat loosely one may say that if multiple durations are available for each v, then duration analysis becomes similar to standard dynamic panel data analysis, where one can get rid of the so-called "fixed effects" before estimating the other parameters. This raises the question to what extent first-differencing of the durations within strata can also be applied to get rid of v. It seems that this is only feasible if the baseline hazard has a particular functional-form specification, notably the Weibull specification. Assume that the duration dependence is described by $\alpha t^{\alpha-1}$ for all spells and strata. In addition, assume that v is the same for all spells in a stratum, and assume for convenience that $\theta_0(x_i) = \exp(x_i'\beta)$. For two spells t_1, t_2 within a stratum, with observed explanatory variables x_1 and x_2, respectively, the difference of equations (26) gives

$$\log t_1 - \log t_2 = -\frac{\beta}{\alpha}(x_1 - x_2) + \frac{\varepsilon_1 - \varepsilon_2}{\alpha}$$

Note that $\varepsilon_1 - \varepsilon_2$ has a fully specified distribution (as the difference of two i.i.d. EV1 random variables). Thus, with Weibull duration dependence, first-differencing results in an equation from which the Weibull parameter and the systematic hazard can be reliably estimated without the need to make any assumption on the unobserved heterogeneity distribution. Indeed, v and x are allowed to be dependent.

The identification results discussed in this subsection have been of enormous importance for applied duration analysis. If two observations are available for each v then the identification of the model does not require an untestable assumption on the tail of the unobserved heterogeneity distribution G anymore, and v and x need not be independent anymore. We only need some fairly innocuous regularity assumptions and normalizations (of course, in addition to proportionality assumptions on the hazard function). The recent applied literature contains a number of studies showing that the estimates of the parameters of interest are robust with respect to the functional-form specification of G, in case of multiple observed durations for each v (see Nielsen et al., 1992, Guo and Rodríguez, 1992, Gönül and Srinivasan, 1993, and Bonnal, Fougère and Sérandon, 1997). These results are in sharp contrast to those found for the single-spell model (Section 5). It should also be noted that Hahn (1994) finds that his result on singularity of the information matrix in the case of single-spell data (see Subsection 5.6) does not carry over to the case of multi-spell data. Moreover, the stratified partial likelihood estimators are $\sqrt{n}-$consistent.

We finish this section by mentioning an important caveat with multi-spell data. This concerns the fact that the analysis of multi-spell data is particularly sensitive to censoring. With single-spell data, many types of censoring are innocuous in the sense that their effect can be captured by standard adjustments to the likelihood function (see Andersen et al., 1993, recall also the discussion in Subsection 4.2). With multi-spell data, one has to be more careful. Consider the case where two durations t_1 and t_2 follow each other in time, and where the data are subject to right-censoring at a fixed duration after the common starting point of the t_1 durations. Then the moment at which t_2 is right-censored is not independent from t_2 itself. To see this, consider individuals for which v is large. For these individuals, t_1 will on average be short. As a result, t_2 will on average start at a relatively early moment. This in turn implies that t_2 will often be right-censored at a relatively high duration. In sum, t_2 and the variable determining the moment at which it is censored are both affected by the unobserved characteristic v. This violates the standard censoring assumptions of duration analysis (see Visser, 1996, for general results, and Keiding, 1998). As a result, standard partial likelihood estimation methods (like the one above) cannot be applied. Moreover, one cannot estimate (characteristics of) the distribution of t_2 in isolation from t_1 (see Ridder and Tunalı, 1999, for an informative exposition). With censoring in general, first-differencing (like above) is not possible. Finally, the value of t_1 may even affect the probability that the beginning of the second spell is observed at all, in which case a subsample of individuals for which both t_1 and t_2 are observed is selective (this is even true if there is no unobserved heterogeneity).[52] Of

[52] In a recent working paper, Woutersen (2000) develops consistent GMM-type estimators that deal with a number of these problems, while treating unobserved heterogeneity as a fixed effect.

course, with censoring, one may still use standard ML estimation methods with random effects. However, if the realization of t_2 is often unobserved then the use of multi-spell data does not provide much gain over the use of single-spell data. In sum, the less censoring in the data, the larger the advantages of multi-spell data.

7 An informal classification of reduced-form multiple-duration models

In general one may think of many different ways to model a relation between duration variables. In the applied econometric literature on the estimation of multiple-duration models, the range of different models is actually not so large. In this section we provide a rather informal model classification that covers most of the models used in practice.[53] The next sections examine the models in more detail. It should be stressed that we are not concerned with abstract point processes where the durations between events can be related for many reasons (see e.g. Snyder and Miller, 1991, for a survey). Also, we are not concerned with the multiple-duration models in engineering where the lifetime of a system depends on the lifetimes of its components. The latter models are often not very useful to describe economic behavior (although they are an important input in economic analyses of machine maintenance; see e.g. Ryu, 1993). As we shall see, some of the models that we consider are more natural when dealing with successive spells in a given state or with successive spells in different states,[54] whereas others are more natural in the case of competing risks, and yet others are useful in all these cases. In fact, the recent empirical literature often uses models that simultaneously allow for two different types of dependence of the duration variables.

[53]See Hougaard (1987) for an older classification, based on statistical model properties.

[54]Again, what constitutes a state depends on the application at hand (i.e. depends on the relevant underlying theoretical framework). It is possible that what in one application are regarded as multiple durations in the same state, are regarded in another application as durations in different states. In practice, for a given individual and a given definition of states, the specifications for the marginal distributions of different spells in a given state are similar, whereas the specifications for the marginal distributions of spells in different states do not contain common parameters or functions.

The MPH model with multi-spell data (Section 6) can also be interpreted as a multiple-duration model, as it specifies the joint distribution of the durations in the spells that an individual experiences. We shall see that this specification is in fact a special case of a popular type of multiple-duration model. For expositional reasons we shall restrict ourselves to two duration variables throughout the remainder of this chapter.

"Lagged" durations

The first popular type of dependence concerns an effect of a realized past duration on the current hazard. This type of dependence was introduced by Heckman and Borjas (1980). Suppose that two durations t_1 and t_2 each follow their own PH model, with $\theta_1(t_1|x_1) = \psi_1(t_1)\theta_{0,1}(x_1)$ and $\theta_2(t_2|t_1, x_2) = \psi_2(t_2)\theta_{0,2}(x_2)\xi(t_1)$, where t_2 starts at or after the moment at which t_1 is realized. Basically, this dependence is modeled by including t_1 as an additional covariate in the hazard for t_2. Usually, the underlying economic theory provides a causal interpretation for this type of dependence.[55] Because of the analogy to a regression model with lagged endogenous variables among the explanatory variables, this dependence is sometimes called "lagged-duration dependence". Obviously, different types of restrictions can be imposed on the model determinants $\theta_{0,1}, \theta_{0,2}, \psi_1$, and ψ_2. For example, if t_1 and t_2 denote durations in the same state then it may be imposed that $\psi_1 \equiv \psi_2$, $x_2 = x_1$, and/or $\theta_{0,2}(x_2) = \theta_{0,1}(x_1)$.

Instead of including the value of t_1 in the individual hazard for t_2, one may also use an indicator of whether the individual has been in the state associated with t_1 during the year before the start of t_2, or indeed any other realization of past behavior. In applied labor economics, these types of dependence have been incorporated in reduced-form models for the effects of labor market programs on subsequent unemployment durations and employment durations. It should be stressed however that these studies also allow for other dependencies; see below for examples.

Recently, in financial econometrics, lagged-duration dependence models have

[55] Here and elsewhere, the relation between the duration variables can be formulated by using the concept of Granger-noncausality. However, for the basic models examined in this chapter, there is no gain from doing this (see Abbring, 1998). See Florens and Fougère (1996) for a formal analysis of causality in more general continuous-time processes.

been used for the analysis of durations between successive market events such as a buy or sell of a security on a stock market (see e.g. Engle and Russell, 1998, and Bauwens and Giot, 1998). In these models, the hazard function of the i^{th} duration depends on the realizations of previous durations by way of an autoregressive scheme. The baseline hazard is assumed to have a Weibull specification with a single common parameter for all durations.

Shocks

The second popular type of dependence concerns situations where two durations occur simultaneously, and where the realization of one duration variable has an immediate effect on the hazard of the other duration variable. This type of dependence has been introduced by Freund (1961). To focus the mind, suppose that the realization of t_1 affects the level of the hazard of t_2 afterwards. This can be captured by the inclusion of an indicator of whether t_1 is realized, as a time-varying regressor in the hazard specification of t_2. For example, the hazard of t_2 can be specified as $\psi_2(t_2)\exp(x_2'\beta_2 + \delta \mathrm{I}(t_1 < t_2))$, where I(.) denotes the indicator function, which is 1 if its argument is true and 0 otherwise. From Subsection 4.2 we know that such a specification requires conditions on the exogeneity of t_1. Basically, t_1 needs to be weakly exogenous, and anticipation by the individual of the future realization of t_1 is ruled out. Note that the individual is allowed to know the (determinants of the) probability distribution of t_1.

The underlying economic theory often provides a causal interpretation for the above type of dependence. Obviously, t_1 and t_2 denote durations in different states, so it does not make sense to impose restrictions across the two hazards.

In practice, it may be too restrictive to assume that the realization of t_1 merely affects the *level* of the hazard of t_2. More generally, the realization may be allowed to affect the whole shape of the hazard of t_2 after the realization of t_1.[56] In applied econometrics, such types of dependence have been incorporated in reduced-form models for the effect of certain treatments[57] on worker labor-market behavior;

[56] In an empirical analysis of panel survey attrition, Van den Berg, Lindeboom and Ridder (1994) examine a slightly different model in which there is a positive *probability* that t_2 is realized immediately after realization of t_1. Here, t_1 and t_2 are the duration until the individual respondent makes a transition to another labor market state, and the duration until attrition from the panel, respectively.

[57] In biostatistics, θ_0 is often called the treatment effect if x captures whether the subject has

we return to this below. In addition, the model described above can be seen as a special case of models in which an individual experiences different stochastic processes which affect each other by way of shifts in the hazard for one process if the other process generates an event. The latter type of models have been used to study the interaction between marital status, number of children, health status, and labor market status. For example, if an unemployed woman marries then her transition rate to employment may drop. It should again be stressed that these studies often also allow for other types of dependence between the duration variables; see below.

Related unobserved determinants

The third type of dependence between duration variables concerns dependence by way of their unobserved determinants. Specifically, consider two durations t_1 and t_2 which each follow their own MPH model, so $\theta_i(t_i|x_i, v_i) = \psi_i(t_i)\theta_{0,i}(x_i)v_i$, with $i = 1, 2$. Then the dependence between t_1 and t_2 given x is modeled by allowing v_1 and v_2 to be related. In Subsection 8.1 below we provide a more precise definition. This multivariate extension to the MPH model is called the Multivariate Mixed Proportional Hazard (MMPH) model. This has in fact been the most popular multiple-duration model by far.[58] Note that the relation between the durations is spurious to the extent that it results from the fact that we do not observe v_i.

The MMPH model applies to cases where the two durations occur simultaneously (possibly with the same starting point) as well as to cases where they occur successively. Again, different types of restrictions can be imposed on the model determinants $\theta_{0,1}, \theta_{0,2}, \psi_1, \psi_2$, and the joint distribution $G(v_1, v_2)$, depending on the extent to which t_1 and t_2 represent durations in the same state. Clearly, the MPH model of Section 6 with a single state and multi-spell data is the special case with $\theta_{0,1} = \theta_{0,2}, \psi_1 = \psi_2$, and $v_1 = v_2$.

The MMPH model is regarded as a convenient and flexible model for dependent durations. Of course, there are often good reasons to suspect the presence of important related unobserved determinants, and by now there is an abundant applied literature in which MMPH models are estimated. In the econometric

received a treatment at the beginning of the spell. Here, we avoid that terminology, and we reserve the term "treatment" for treatments occurring during a spell.

[58] Flinn and Heckman (1982b) provide an early analysis of this model.

contributions to this literature, the variety of types of states and durations that are considered is vast. Flinn and Heckman (1982b, 1983), Coleman (1990), and Rosholm (1997) estimate MMPH models for the durations of unemployment, employment, etc. in order to study transition rates between different labor market states. Generally, the unobserved determinants of the durations spent in different states are allowed to be related, and the unobserved determinants of different durations spent by an individual in the same state are assumed to be identical. In their studies of attrition in longitudinal panel survey data, Van den Berg, Lindeboom and Ridder (1994), Carling and Jacobson (1995) and Van den Berg and Lindeboom (1998) estimate MMPH models for the joint durations of labor-market spells (like a spell of unemployment or a job spell) and the duration of panel survey participation. Lillard and Panis (1998) include attrition in a similar way in their model for the joint durations of marriage, non-marriage, and life. Note that this approach to attrition is in line with the popular modeling setup for sample selection introduced by Heckman (1979).

As we saw in Section 6, MPH models are sometimes estimated under the assumption that the unobserved heterogeneity term is identical across different physical individuals within some group or stratum. Sastry (1997) extends this setup by allowing each individual to belong to two groups with different aggregation levels (families and towns). There is unobserved heterogeneity across each type of group. This effectively amounts to an MMPH specification for the durations of members of different families living in the same town. Similarly, the approach in studies on lifetime durations where the unobserved heterogeneity terms are assumed to be identical across siblings can be generalized to allow v_1 and v_2 for siblings to be a sum of a common determinant and an independent person-specific component (see e.g. Petersen, 1996, Yashin and Iachine, 1997, and Zahl, 1997, for applications).[59] Such a specification for G has gained less popularity in econometrics, for the obvious reason that in econometric applications the association of unobserved heterogeneity to genetic factors is less compelling.

[59]The applications of this paragraph illustrate a disadvantage of the "multi-state / multi-spell" terminology: sometimes two spells are in the same state but one does not want to impose that the unobserved heterogeneity terms are identical, so that the multi-spell setup of Section 6 does not apply.

Combinations of dependencies

The presence of related unobserved determinants is particularly important if one is interested in one of the other two types of dependence that we described above. The estimate of the causal effect will be biased if one ignores the spurious dependence that results from the related unobserved determinants. To deal with this, the empirical model should take account of this spurious dependence. The model should allow both for a causal effect *and* for related unobserved heterogeneity.

As examples of a combination of lagged duration dependence and related unobserved heterogeneity, see Heckman, Hotz and Walker (1985), who allow "lagged" durations between the births of previous children to affect the hazard of the duration of the current birth interval, and who allow for correlated unobserved heterogeneity as well (see Omori, 1997, and Lancaster, 2000b, for other examples). Lillard (1993), Lillard and Panis (1996), Abbring, Van den Berg and Van Ours (1997), Eberwein, Ham and LaLonde (1997), and Van den Berg, Van der Klaauw and Van Ours (1998) analyze models where the realization of one duration variable has an immediate effect on the hazard of the other duration variable, allowing for related unobserved heterogeneity in order to deal with selectivity. Let us examine them in somewhat more detail. Abbring, Van den Berg and Van Ours (1997) and Van den Berg, Van der Klaauw and Van Ours (1998) study the effect on the exit rate out of unemployment of a punishment for insufficient search effort. The duration until punishment is modeled by way of an MPH model, and the exit rate out of unemployment permanently shifts to another level at the moment the punishment is applied. Lillard (1993) estimates a model for the joint durations of marriage and time until conception of a child, and his model allows the rate at which the marriage dissolves to shift to another level at moments of child birth. Lillard and Panis (1996) estimate a model on the joint durations of marriage, non-marriage, and life, and their model allows the death rate to shift to another level at moments of marriage formation and dissolution. Eberwein, Ham and LaLonde (1997) estimate a (discrete-time) model for the effect of participation in training programs on individual labor market transitions, and they allow the exit rate out of unemployment to shift to another level at the moment of inflow into the program. See Van den Berg, Holm and Van Ours (1999) for a similar analysis in continuous time. In all these applications, the duration variable t_1 needs to satisfy the exogeneity conditions of Subsection 4.2 for *given* values of the unobserved heterogeneity terms. This rules out anticipations of the realizations of t_1, but the individual is allowed to know the (determinants of the) probability distribution of t_1.

In the applied literature on the effects of training on unemployment durations, "training" is often regarded to be a separate labor market state, and the effect of training on subsequent labor market transitions can then be captured by a model with lagged-duration dependence (or a model where the fact that one has had any training is allowed to affect subsequent transitions). In order to deal with selectivity of those who enrol in training, it is important to allow for related unobserved heterogeneity terms affecting the inflow into training as well as the other transition rates. Gritz (1993) and Bonnal, Fougère and Sérandon (1997) contain sophisticated examples of such analyses. Ham and LaLonde (1996) use experimental data to estimate models for the effects of training on individual labor market transition rates.

In the *absence* of unobserved heterogeneity, the specification, identification, and ML estimation of models with lagged-duration dependence is relatively straightforward. The same holds for models with changes in the hazard of one duration in response to realization of the other duration (given appropriate assumptions on the direction of the causality; see Florens and Fougère, 1996). However, models with related unobserved heterogeneity terms are less transparent. In the next section we therefore examine MMPH models in detail. Subsequently, in Section 9, we briefly examine the models where related unobserved heterogeneity is combined with a "causal" effect of one duration on the other (that is, we examine a combination of lagged duration dependence and unobserved heterogeneity, and a combination of a shift in the hazard and unobserved heterogeneity).

Some theoretical considerations

We finish this section by stressing that, like in Section 4, it is often not clear to what extent the reduced-form specifications of the dependence between two durations can be justified by economic-theoretical models. This is particularly true for models where the hazard of one duration immediately changes in response to the realization of the other duration. In many cases, individuals may anticipate the realization of the other duration, and the moment at which the anticipation starts is often unobserved. In applications this has to be examined carefully.

In the analysis of MMPH models, as a rule, the assumed parametric family of the joint unobserved heterogeneity distribution $G(v_1, v_2)$ treats v_1 and v_2 in a symmetric way: given the unknown parameters of G, the role of v_1 and v_2 in $G(v_1, v_2)$ can be interchanged without changing G. In particular, if G is continuous then the supports of v_1 and v_2 are assumed to be the same, and if G is discrete then the numbers of points of support are assumed to be the same for v_1 and v_2. It is sometimes difficult to justify such symmetric distributions with economic theory. If, according to the theory, individuals improve their situation when ending one spell and starting another, then the characteristics associated with the second spell should be "superior" in some sense to those of the first spell. If v_1 represents the characteristics of the first spell and v_2 of the second, then this suggests that the support of v_2 should depend on the realization of v_1. Consider for example the on-the-job search model discussed in Subsection 5.5.3. If one observes two consecutive job spells and if the wages are unobserved, then the unobserved heterogeneity term of the second spell exceeds the term of the first spell. Unfortunately, such bivariate heterogeneity distributions have not yet been studied (see Koning et al., 2000, for an application in a structural analysis of an on-the-job search model).

Finally, we address whether the hazards of different durations of the same individual depend on the same set of explanatory variables or not. Economic theory often predicts that both hazards depend on the individual's behavior, and that the forward-looking individual's optimal strategy depends on all structural determinants. For example, in a job search model with two possible employment destination states, the decision on whether to accept a job offer depends on the arrival rates and wage offer distributions of both types of employment, regardless of the employment type of the actual offer (see Thomas, 1998). In such cases, if the observed explanatory variables are characteristics of the individual himself, then it does not make sense to exclude elements of x from one hazard that are included in the other hazard. In other words, in such cases, $x_1 = x_2$ (note incidentally that this provides an argument against the assumption that unobserved heterogeneity is independent across spells for a given individual; see also Lillard, 1993). In the event that the researcher observes a determinant of one of the hazards whereas this determinant is assumed to be unobserved by the individual, then it makes sense to include this determinant only in the corresponding hazard. Finally, if one hazard is mechanical and independent of the individual's behavior then obviously it does not need to depend on the determinants of the other hazard (see Van den Berg, 1990b, and Ryu, 1993, for examples).

8 The Multivariate Mixed Proportional Hazard model

8.1 Definition

In this subsection we define the MMPH model. Next, Subsection 8.2 deals with identification of this model under different situations with respect to the timing of the two underlying spells. We assume that the situation is either such that both durations always start at exactly the same point of time, or that one duration necessarily follows the other. In Subsection 8.3 we discuss parametric specifications for the joint distribution of unobserved heterogeneity and the degree of flexibility of the corresponding models.

For the sake of convenience, we again use the term "individual" to denote the subject that experiences certain spells. In the first situation with respect to the timing of the spells (starting at the same time) we consider the population of individuals in the inflow into the states corresponding to the duration variables, whereas in the second situation (successive durations) we consider the population of individuals in the inflow in the state corresponding to the first duration. Flinn and Heckman (1982b), Chesher and Lancaster (1983), and Ham and LaLonde (1996) consider less "ideal" sampling designs.

We assume that all individual differences in the hazard function of t_1 can be characterized by observed explanatory variables x and unobserved characteristics v_1. Similarly, all individual differences in the hazard function of t_2 can be characterized by observed explanatory variables x and unobserved characteristics v_2. (Of course, one may impose exclusion restrictions on the set of elements of x that is allowed to affect the systematic hazard $\theta_{0,i}(x)$ associated with exit i.) For an individual with explanatory variables x, v_1, v_2, the hazard functions of t_1 and t_2 conditional on x, v_1, v_2 are denoted by $\theta_1(t_1|x, v_1)$ and $\theta_2(t_2|x, v_2)$. The MMPH model is now defined by

Definition 2 : MMPH model. *There are functions $\psi_1, \psi_2, \theta_{0,1}, \theta_{0,2}$ such that for every t_1, t_2, x, v_1, v_2 there holds that*

$$\theta_1(t_1|x,v_1) = \psi_1(t_1) \cdot \theta_{0,1}(x) \cdot v_1$$
$$\theta_2(t_2|x,v_2) = \psi_2(t_2) \cdot \theta_{0,2}(x) \cdot v_2 \tag{29}$$

For convenience, we take $\psi_1, \psi_2, \theta_{0,1}, \theta_{0,2}, v_1, v_2$, and the distribution G of v_1, v_2 in the population to satisfy the regularity assumptions that correspond to Assumptions 1–4 for ψ, θ_0, v, G in the MPH model.

Conditional on x, v_1, v_2, the durations t_1 and t_2 are independent. Conditional on x, the variables t_1 and t_2 are only dependent if v_1 and v_2 are dependent. So, in the case of independence of v_1 and v_2, the model reduces to two unrelated ordinary MPH models for t_1 and t_2.

In terms of a regression specification with $\theta_{0,i}(x) = \exp(x'\beta_i)$, this model can be rewritten as

$$\log \int_0^{t_1} \psi_1(u)du = -x'\beta_1 - \log v_1 + \varepsilon_1$$
$$\log \int_0^{t_2} \psi_2(u)du = -x'\beta_2 - \log v_2 + \varepsilon_2 \tag{30}$$

where ε_1 and ε_2 are i.i.d. EV1 distributed, but where v_1 and v_2 may be related.

Now consider the joint distribution of t_1 and t_2 given x. The joint density $f(t_1, t_2|x)$ can be expressed as

$$f(t_1, t_2|x) = \int_0^\infty \int_0^\infty f_1(t_1|x, v_1) f_2(t_2|x, v_2) \quad dG(v_1, v_2)$$

in which we already implicitly assume that v_1, v_2 are independent of x, and in which the probability density function of $t_i|x, v_i$ is for convenience denoted by $f_i(t_i|x, v_i)$. The latter density can of course be expressed in terms of the determinants of θ_i (see Section 2). Let $z_i(t_i)$ denote the integrated baseline hazard associated with t_i. The joint survivor function of t_1 and t_2 can then be expressed as

$$\overline{F}(t_1, t_2|x) = \int_0^\infty e^{-z_1(t_1)\theta_{0,1}(x)v_1 - z_2(t_2)\theta_{0,2}(x)v_2} dG(v_1, v_2)$$

In many applications, the individual likelihood contribution is based on the density above (that is, if the unobserved heterogeneity terms are not treated as incidental parameters). In terms of panel data analysis, this means that v_1, v_2 are treated as "random effects" when estimating the model with Maximum Likelihood.

8.2 Identification results

In this subsection we consider identification results for the MMPH model. It is important to stress that no parametric functional form assumptions are made on the underlying functions $\theta_{0,i}, \psi_i$ and G, so, as in Subsection 5.2, we are concerned with *nonparametric* identification.

8.2.1 Competing risks

Recall from Subsection 8.1 that we consider two different situations with respect to the timing of the two spells. In the first situation, both spells start at the same point of time for a given individual, and the individual is observed until the first duration is completed. This is called a competing-risks model, as one may envisage the individual having two options to leave the current state, and the realization of one option is necessary and sufficient for leaving the state. In the second situation with respect to the timing of the spells, the two spells cannot overlap. Moreover, in the second situation both durations can be followed until completion, so there is more information available than in the first situation (see Subsection 8.2.2 below).

In the competing-risks setting, the data provide information on $\min\{t_1, t_2\}$ and on $\arg\min_i t_i$ (i.e. on which duration is the one that ends first). So, assume that the data provide the distribution of this "identified minimum". It is well known that this does not suffice to identify the most general competing-risks model (with an arbitrary joint distribution for t_1, t_2, without covariates). In particular, for every model with dependent t_1, t_2 there is an observationally equivalent model with independent t_1, t_2 (see e.g. Lancaster, 1990).

Now let us assume that t_1 and t_2 are generated by an MMPH model with regularity assumptions corresponding to Assumptions 1–4. As in Subsection 5.2, some additional assumptions are needed for identification. These include the equivalents of Assumption 5 (so x is independent of v_1, v_2), Assumption 7 (normalizations), and Assumption 8 ($\mathrm{E}(v_i) < \infty$). In addition, we need to strengthen Assumption 6 on the dispersion of x,

Assumption 9 : Variation in observed explanatory variables in the competing-risks setting. *The functions $\theta_{0,1}(x), \theta_{0,2}(x)$ attain all values in a set $(0, \overline{\theta}_{0,1}) \times (0, \overline{\theta}_{0,2})$ with $0 < \overline{\theta}_{0,1}, \overline{\theta}_{0,2}$, when x varies over the set $\mathcal{X}$ of possible values of x.*

If $\theta_{0,i}(x) = \exp(x'\beta_i)$ then sufficient for this is that x has *two* continuous covariates which affect both hazards θ_i but with different coefficients for different i, and which are not perfectly collinear. Moreover, in the population, these covariates must attain all values ranging to minus infinity.

Heckman and Honoré (1989) prove the nonparametric identification of the model under these assumptions. In fact, they strengthen Assumption 9 by taking $\overline{\theta}_{0,i} = \infty$, because they examine a class of models that is somewhat more general than the class of MMPH models (see Abbring and Van den Berg, 2000b). In any case, note that Assumption 9 is stronger than Assumption 6 on the range of values that θ_0 attains in the MPH model. This is not surprising. However, it is important to note that the identification does not require exclusion restrictions on the hazard specification of either duration. Moreover, identification does not require parametric functional form restrictions on the distribution of unobserved heterogeneity. In the case of binary data on the "identified minimum" (i.e., it is observed which duration ends first but not when) such restrictions are necessary to achieve identification. This illustrates the fact that the timing of events in duration data provides a valuable source of information concerning the underlying model.

It is interesting to obtain some insight into the identification of whether the durations are dependent or not, since this distinguishes the above identification result from the earlier literature in which competing risks models without covariates were examined. In the sequel of this subsection we use T_1, T_2 to denote the random duration variables, and t_1, t_2 to denote realizations of these. We define

$$\theta_1^*(t_1 | x, T_2 > t_1)$$

to be the hazard of the duration T_1 at the value t_1, conditional on x and conditional on the duration T_2 exceeding t_1. More generally, the hazard $\theta_1^*(t_1 | x, T_2 > t_2)$

corresponds to the conditional distribution of $T_1|x, T_2 > t_2$. We evaluate this hazard for given t_1 and t_2, and in fact we take $t_2 = t_1$. Obviously, the hazard $\theta_2^*(t_2|x, T_1 > t_2)$ can be defined analogically. It is important that the "conditional" hazards $\theta_1^*(t_1|x, T_2 > t_1)$ and $\theta_2^*(t_2|x, T_1 > t_2)$ are observable quantities, as they can be expressed in terms of the distribution of the data. (Note that the "marginal" hazards $\theta_i(t_i|x)$ are unobserved due to the competing risks setting.)

If v_1 and v_2 are independent, then

$$\theta_1^*(t_1|x, T_2 > t_1) = \theta_1(t_1|x) \quad \text{and} \quad \theta_2^*(t_2|x, T_1 > t_2) = \theta_2(t_2|x)$$

The assumption in Heckman and Honoré (1989) on the values that can be attained by $\theta_{0,i}(x)$ implies that $\theta_{0,1}(x)$ and $\theta_{0,2}(x)$ are not perfectly related, and that there is some independent variation in both. As a result, if v_1 and v_2 are independent then $\theta_{0,2}(x)$ does not affect $\theta_1^*(t_1|x, T_2 > t_1)$, and $\theta_{0,1}(x)$ does not affect $\theta_2^*(t_2|x, T_1 > t_2)$.

Now let us examine what happens if v_1 and v_2 are dependent. It is straightforward to show that

$$\theta_1^*(t_1|x, T_2 > t_1) = \frac{\mathrm{E}_v \left[\theta_1(t_1|x, v_1) \exp\left(-\int_0^{t_1} \theta_1(u|x, v_1)du - \int_0^{t_1} \theta_2(u|x, v_2)du\right)\right]}{\mathrm{E}_v \left[\exp\left(-\int_0^{t_1} \theta_1(u|x, v_1)du - \int_0^{t_1} \theta_2(u|x, v_2)du\right)\right]}$$

with θ_i as in (29), and with E_v denoting the expectation with respect to the bivariate distribution $G(v_1, v_2)$. If we differentiate this with respect to $\theta_{0,2}(x)$ then the resulting expression has the same sign as

$$-\mathrm{Cov}(v_1, v_2|x, T_1 > t_1, T_2 > t_1)$$

(provided that $t_1 > 0$). If v_1 and v_2 are dependent then in general there are many values of t_1 such that the above expression is nonzero.

In sum, the derivative of $\theta_1^*(t_1|x, T_2 > t_1)$ with respect to $\theta_{0,2}(x)$ and its mirror image for t_2 are informative on the dependence or independence of the unobserved heterogeneity terms. This is intuitively very plausible. If the systematic hazard of t_2 does not directly affect the individual hazard of t_1 but does affect the observed hazard of t_1 then this indicates that there is a spurious relation between the durations by way of their unobserved determinants. It should again be stressed that this is not based on an exclusion restriction in the usual sense of the word. All explanatory variables are allowed to affect (the means of) both duration variables

– they are just not allowed to affect the whole duration distributions in the same way.[60]

The above results are based on the availability of "single-spell" data. In the present context, this means that for each individual in the sample there is one observation of the "identified minimum" (which consists of $\min\{t_1, t_2\}$ and $\arg\min_i t_i$). Now suppose that the individual-specific value of the v_1, v_2 pair is invariant over time. In a recent working paper, Abbring and Van den Berg (2000b) show that some of the assumptions made by Heckman and Honoré (1989) can be weakened substantially if the data provide multiple observations on the identified minimum for each individual.

8.2.2 Successive durations

If the two spells are successive, and both durations can be followed until completion, then the data provide the joint distribution $F(t_1, t_2|x)$. In fact, it is merely for expositional reasons that we take the spells to be successive: if they occur (partly) simultaneously and are both observed until completion then the results of this subsection are valid as well, provided that the durations satisfy the model as defined in Subsection 8.1.

The most general model specification does not impose restrictions across the marginal duration distributions, so it allows for $\psi_1 \neq \psi_2, \theta_{0,1} \neq \theta_{0,2}$, and $v_1 \neq v_2$. For both marginal hazard functions in this model we make regularity assumptions corresponding to Assumptions 1–4. In addition, we adopt the equivalents of the Assumptions 5–8 that were made to identify the MPH model. Honoré (1993) shows that under these assumptions the MMPH model is identified. (Assumptions 6 and 8 may be jointly replaced by Assumptions 6b and 8b.)

This result is not surprising, because the data on $t_i|x$ identify the determinants of the MPH model for t_i (which are $\psi_i, \theta_{0,i}$, and the marginal distribution of v_i), provided that the assumptions for identification of this MPH model are satisfied. The relation between v_1 and v_2 is subsequently identified from the observed relation between t_1 and t_2 given x.

Sometimes it makes sense to impose a priori restrictions across the mar-

[60] Of course, the $\theta_{0,i}(x)$ are not directly observed. Heckman and Honoré (1989) identify these by examining data at zero durations. Whether this can be used to construct a useful test statistic on independence remains to be seen.

ginal duration distributions. The most restrictive specification imposes that $\psi_1 = \psi_2, \theta_{0,1} = \theta_{0,2}$, and $v_1 = v_2$. We already know from Section 6 that this model is identified under weak assumptions. Now let us consider an intermediate case in which we impose that $v_1 = v_2$ but allow the baseline hazards ψ_1 and ψ_2 to be different. In addition, we do not assume that there are observed explanatory variables x. In other words, the analysis is conditional on a given value of x, and we allow for full interaction of the actual value of x with the model determinants: ψ_i may depend on x in an unspecified way, and v and x may be dependent in the population (from this point of view we do not consider an "intermediate" case, as this generalizes the MMPH specification). Thus,

$$\theta_i(t|x, v) = \psi_i(t|x) \cdot v, \qquad v|x \sim G(v|x)$$

This includes of course as a special case that $\psi_i(t|x)$ can be written as $\psi_i(t)\theta_{0,i}(x)$. We make regularity assumptions corresponding to Assumptions 2–4. Honoré (1993) shows that this model is identified, provided that a normalization is imposed on the integrated baseline hazard (analogical to Assumption 7). Note that we do not need to make assumptions corresponding to the previously made Assumptions 5, 6, and 8. Perhaps the most important issue here is that identification does not require independence of v and x. In many applications, such independence is difficult to justify. Like in Section 6, if unobserved heterogeneity values are identical across different durations then the model is similar to a standard dynamic panel data model.

8.3 Specification of the bivariate unobserved heterogeneity distribution

8.3.1 Dimensionality

The types of justifications used for parametric functional forms of G in MPH models are often unavailable for MMPH models. This is particularly true for the choice of a specification for the dependence of v_1 and v_2. In this subsection we focus on the choice of the dimensionality of the distribution of G (or more accurately, the dimension of the support of G). In Subsection 8.3.2 we then examine the types of dependence that can be generated by different parametric functional forms for a G with a given dimensionality.

The so-called "one-factor loading specification" has been a popular specification for a bivariate distribution of unobserved heterogeneity terms in MMPH models (see Flinn and Heckman, 1982b, 1983, for early applications, and Heckman, Hotz and Walker, 1985, Heckman and Walker, 1987, 1990, and Bonnal, Fougère and Sérandon, 1997, for subsequent applications). This specification reduces the dimensionality of the distribution G from 2 to 1. In particular, it assumes that there is a univariate random variable z such that

$$v_i = \exp(\alpha_i + \gamma_i z) \qquad i = 1, 2 \tag{31}$$

(Note that this z does not refer to the integrated baseline hazard here.) This specification can be straightforwardly generalized to a higher number of different durations as well as a higher dimension of the random variable z. If z is two-dimensional then we obtain a "two-factor loading specification", etc.

The two (related) advantages of the "factor loading specifications" are (1) they restrict the number of unknown parameters, leading to a sparse specification, and (2) they limit the computational burden of the estimation of the model. The number of parameters related to G equals the number of parameters of the distribution of z, plus the number of α_i and γ_i parameters, minus normalizations. This typically increases linearly with the number of different durations n. If $v_1, \ldots, v_n$ has a genuine multivariate distribution then the number of parameters related to G typically increases quadratically with n. To illustrate the computational advantage, consider the case where $\log v_1, \ldots, \log v_n$ has a multivariate normal distribution. The evaluation of the joint density function of $t_1, \ldots, t_n$ then requires the evaluation of an $n-$dimensional integral. However, if the v_i are related by a one-factor loading specification then the integral is one-dimensional. See for example Bonnal, Fougère and Sérandon (1997), where $n = 8$. Note that computational burden is less of a problem in the case of discrete v_i and n smaller than, say, 4.

Hougaard (1987) stresses that it is too restrictive to assume that $v_1 \equiv v_2$ if the corresponding spells do not concern the same state. If (i) $v_1 \equiv v_2$, and (ii) both durations are always observed, and (iii) each duration is described by an identified MPH model, then the full unobserved heterogeneity distribution is completely identified from data on only one of the durations. We now show that somewhat similar problems may arise in the case of a one-factor loading specification for G.

Indeed, the main disadvantage of the one-factor loading specification concerns the relation it imposes on the marginal duration distributions on the one hand, and the dependence of the durations on the other. If $\text{Var}(v_1) > 0$ and $\text{Var}(v_2) > 0$ then it automatically follows that $\text{Cov}(v_1, v_2) \neq 0$. So if the data provide evidence for unobserved heterogeneity in the marginal distributions of t_1 and t_2, then the model implies that these durations must be dependent. Similarly, if the durations are independent, then the model implies that there is no unobserved heterogeneity for at least one of the durations. If the dependence between the durations changes, then necessarily the marginal duration distributions change as well. Lindeboom and Van den Berg (1994) show in detail that these may amount to serious restrictions on the specification of the full model.

To illustrate this issue, suppose that the distribution of z belongs to a parametric family of distributions with two parameters: a location parameter μ and a scale parameter σ (for example, z has a normal distribution with parameters μ and σ). Then

$$z = \mu + \sigma \tilde{z},$$

where $\tilde{z}$ has a completely specified distribution. By substituting this into (31), it is clear that we can only identify $\alpha_1 + \gamma_1\mu$, $\alpha_2 + \gamma_2\mu$, $\gamma_1\sigma$, and $\gamma_2\sigma$. This implies that in effect we only have two parameters at our disposal to capture the 3 second moments of $\log v_1, \log v_2$ (which are $\text{Var}(\log v_1)$, $\text{Var}(\log v_2)$, and $\text{Cov}(\log v_1, \log v_2)$).

8.3.2 The dependence between the durations

In this subsection we examine the dependence of the two duration variables in the MMPH model. For this purpose we use some summary measures of the association between two random variables. For a given association measure we focus on two issues: first, which range of values of this association measure can be attained by the MMPH model in general, and secondly, to what extent is this range further narrowed if G is assumed to belong to specific families of distributions. The first issue is of importance for a comparison of the MMPH model to other models for the dependence between duration variables. The second issue is of importance for a comparison of the flexibility of different families of heterogeneity distributions, and to obtain insight into the range of bivariate models that can be generated by a specific G. The results in this subsection are from Van den Berg (1997).

The regression-type specification of the MMPH model (see equation (30)) suggests that $\text{Corr}(\log z_1(t_1), \log z_2(t_2)|x)$ may be an interesting summary measure of the association between t_1 and t_2. Unfortunately it turns out that for our purposes it is not, because it can attain every value in $(-1, 1)$ for given baseline hazards, by choosing an appropriate G. Moreover, it can attain every value in $(-1, 1)$ within the popular parametric families of distributions for G. Consider instead $\text{Corr}(t_1, t_2|x)$, and assume for the moment that the baseline hazards are constant. The correlation of the duration variables is informative on the strength of the linear relationship between these variables. It is a commonly used measure that is readily understood. Here, it equals

$$\text{Corr}(t_1, t_2|x) = \frac{\text{Cov}(\frac{1}{v_1}, \frac{1}{v_2})}{\prod_{i=1}^{2} \left[\text{Var}(\frac{1}{v_i}) + \text{E}(\frac{1}{v_i^2})\right]^{1/2}} \tag{32}$$

Note that it does not depend on x and that its sign equals the sign of $\text{Corr}(1/v_1, 1/v_2)$.

Van den Berg (1997) shows that

$$-\frac{1}{3} < \text{Corr}(t_1, t_2|x) < \frac{1}{2}$$

regardless of the values of $\theta_{0,1}(x)$ and $\theta_{0,2}(x)$, and regardless of the shape of $G(v_1, v_2)$ (but provided that the right-hand side of (32) exists). The inequalities are sharp in the sense that they can be approached arbitrarily closely by choosing appropriate G.

The result above (and most of the results below) can be easily generalized to models with Weibull baseline hazards. In that case, the upper and lower bound depend on the parameters of the baseline hazard, but they are always strictly between -1 and 1, and the lower bound is always closer to zero than the upper bound.[61]

In the empirical literature, the most frequently used families of distributions for v_1, v_2 are (1) the family of bivariate discrete distributions with two points of

[61] Similar results can be derived for bivariate accelerated failure time models and bivariate duration models in discrete time, notably the discretized (i.e. rounded-off) bivariate MPH model and the rather popular bivariate discrete-time duration model in which the exit probabilities have logistic specifications.

support for v_1 and for v_2, and (2) the family of bivariate normal distributions for $\log v_1, \log v_2$. These families include as special cases the one-dimensional distributions with perfect correlations (these can be represented by the one-factor loading specification (31)). Coleman (1990), Van den Berg, Lindeboom and Ridder (1994), Carling and Jacobson (1995), and Van den Berg and Lindeboom (1998) adopt multivariate discrete distributions for G,[62], whereas Butler, Anderson and Burkhauser (1986), Lillard (1993), Xue and Brookmeyer (1996), Lillard and Panis (1996, 1998), and Ng and Cook (1997) adopt multivariate normal distributions.[63] It turns out that in the discrete case, every value in $(-1/3, 1/2)$ can be attained. By implication, this is also true in the case of more than two points of support for each v_i. In the normal case, $\text{Corr}(t_1, t_2|x)$ can only attain values in $[-3+2\sqrt{2}, 1/2)$, where the lower bound equals about -0.17.

The lower bound $-1/3$ is attained for a discrete distribution for v_1, v_2 such that $\Pr(v_1 = c_1, v_2 = \infty) = \Pr(v_1 = \infty, v_2 = c_2) = 1/2$, with $0 < c_1, c_2 < \infty$.[64] In that case, the bivariate distribution of $t_1, t_2|x$ is such that, with probability $1/2$, $t_1|x$ is zero and $t_2|x$ has an exponential distribution, and with probability $1/2$ this holds with t_1 and t_2 interchanged. We conclude that in an MMPH model these (and similar) duration distributions cannot be generated if $\log v_1, \log v_2$ has a normal distribution, which may be a disadvantage of the latter if one is interested in a flexible specification.[65]

[62]Engberg, Gottschalk and Wolf (1990) estimate a bivariate discrete-time duration model in which the individual per-period exit probabilities are logistic functions of $\psi_i(t_i)\theta_{0,i}(x)v_i$, and in which G has a bivariate discrete distribution. Meghir and Whitehouse (1997) estimate a similar discrete-time model, with a genuine bivariate discrete distribution, but with probit specifications for the exit probabilities. Heckman, Hotz and Walker (1985), Heckman and Walker (1987, 1990) and Gritz (1993) adopt discrete distributions for z in a one-factor loading specification. Card and Sullivan (1988), Mroz and Weir (1990), Ham and LaLonde (1996) and Eberwein, Ham and LaLonde (1997) estimate discrete-time bivariate duration models with logistic probabilities and a one-factor loading specification for z with a discrete distribution.

[63]Flinn and Heckman (1982b, 1983) and Bonnal, Fougère and Sérandon (1997) adopt normal distributions for z in a one-factor loading specification. In a sensitivity analysis, the latter study also adopts a discrete distribution for z.

[64]This should not be interpreted as an advantage of discrete random variables for v_1, v_2 vis-à-vis continuous random variables, for one can construct families of bimodal continuous distributions for G such that $-1/3$ can be approached arbitrarily closely.

[65]Butler, Anderson and Burkhauser (1989) assume v_1, v_2 to have a bivariate discrete distribution with points of support that are fixed in advance. This means that the only parameters

For the general model as well as within the parametric families discussed above, the distributions that give the largest and smallest possible value of $\text{Corr}(t_1, t_2|x)$ are such that $\log v_1$ and $\log v_2$ are perfectly correlated. This means that the range of values for $\text{Corr}(t_1, t_2|x)$ is the same as in the case of a one-factor loading model (see equation (31)) with an appropriate distribution of z. In other words, a reduction of the class of G to one-factor loading specifications does not further restrict the range of values that $\text{Corr}(t_1, t_2|x)$ can attain.[66] From this point of view, one-dimensional random variation in the unobserved heterogeneity terms is sufficient for maximum flexibility in terms of the correlation of the durations.

As an alternative measure of association, consider Kendall's τ (or "Kendall's coefficient of concordance"). This is the most popular global ordinal measure of association in the literature on multivariate durations (see e.g. Genest and MacKay, 1986, Oakes, 1989, and Guo and Rodríguez, 1992). There are several equivalent ways to formally define it. The definition given by Kendall (1962) is particularly useful for general multivariate duration models,

$$\tau(t_1, t_2|x) = 4\text{E}(F(t_1, t_2|x)) - 1$$

where the expectation is taken with respect to $F(t_1, t_2|x)$ itself. Kendall's τ only attains values in $[-1, 1]$. It is an ordinal measure, and it is informative on the strength of any monotone relation. It equals 1 (-1) if and only if t_2 is a monotone increasing (decreasing) function of t_1. Because it is invariant under monotone transformations of the random variables, the value of $\tau(t_1, t_2|x)$ in the MMPH model does not depend of the baseline hazards or on the values of the systematic hazards (so the baseline hazards can be taken as constants, and the conditioning on x can be omitted). As a result, it only depends on the distribution G of the unobserved heterogeneity terms, which is exactly the part of the model that causes the dependence of the durations.

For convenience, assume that $G(v_1, v_2)$ follows a one-factor loading specification, i.e. suppose (31) holds. It turns out that all values between -1 and 1 can be attained by $\tau(t_1, t_2)$, within any family of continuous distributions for z. However,

of G to be estimated are the probabilities associated with these points of support. This can be shown to narrow the range of values of $\text{Corr}(t_1, t_2|x)$ as well, in particular if the points for v_1 or v_2 are chosen to be relatively close to one another (see Van den Berg, 1997, for examples).

[66] Note that if $v_1 \equiv v_2$ then this range reduces to $(0, 1/2)$.

if z (and therefore v_i) is restricted to have a discrete distribution with n points of support ($n = 2, 3, \ldots, \infty$), then

$$-1 + \frac{1}{n} < \tau(t_1, t_2) < 1 - \frac{1}{n}$$

These inequalities are sharp in the sense that they are approached arbitrarily closely for appropriate values of the parameters in the one-factor loading specification (31).

The results for τ are clearly quite different from those for the correlation coefficient. This is because τ detects linear and nonlinear monotone relations alike, and it does not depend on the relative magnitudes of the duration variables, but only on their ordering. The fact that the range of values of $\tau(t_1, t_2)$ is restricted for discrete distributions with finite n can be explained as follows. In this case, the population can be subdivided into a finite number of groups of individuals, and within these groups, all individuals are the same in terms of their v_1 and v_2. This implies that there is a positive probability that two random drawings of t_1 and t_2 are from the same group. Now consider all observations for a single group. Because they all have the same v_1 and v_2, there is no relation at all between t_1 and t_2 within the group. This restricts the population value of $\tau(t_1, t_2)$. It does not affect the range of values of $\text{Corr}(t_1, t_2|x)$ because the "within-group" lack of correlation can be made quantitatively unimportant by making the "between-group" differences large.

In all cases, the bounds for $\tau(t_1, t_2)$ are attained by "spreading out" the heterogeneity distribution as much as possible. If z is continuous then the resulting bivariate distribution of $t_1, t_2|x$ is such that all probability mass is on a single curve for t_1 and t_2. We conclude that in an MMPH model such a duration distribution cannot be generated if z has a discrete distribution with a finite number of points of support. This suggests that it is useful in empirical applications to try to increase the number of mass points.

We finish this subsection by noting that in applications it may also be interesting to examine the dependence of the residual duration variables if one conditions on survival up to a certain duration. It may also be interesting to examine how the (non-causal) effect of the realization of one duration variable on the hazard rate of the other changes with the realized value of the first duration variable. Oakes (1989), Anderson et al. (1992), Hougaard, Harvald and Holm (1992b) and

Yashin and Iachine (1999) provide analyses for the general case, and they also discuss how the dependence patterns are affected by the functional form of G.

9 Causal duration effects and selectivity

9.1 Lagged endogenous durations

In this subsection we briefly examine bivariate duration models with lagged-duration dependence as well as mutually related unobserved heterogeneity terms. Recall from Section 7 that such models have been used to study the impact of the length of an unemployment spell on the length of the next unemployment spell. Also recall that the estimate of the effect of the previous duration is biased if one ignores the spurious dependence from related unobserved determinants.

In terms of the hazards, the model specification reads

$$\begin{aligned} \theta_1(t_1|x,v_1) &= \psi_1(t_1)\cdot\theta_{0,1}(x)\cdot v_1 \\ \theta_2(t_2|t_1,x,v_2) &= \psi_2(t_2)\cdot\theta_{0,2}(x)\cdot\xi(t_1)\cdot v_2 \end{aligned} \tag{33}$$

and we make the following regularity assumption on the function ξ,

Assumption 10 *The function $\xi(t)$ is positive for every $t\in[0,\infty)$.*

If v_1 and v_2 are independent, then, conditional on x, the durations t_1 and t_2 are only dependent if $\xi(t_1)$ is not a constant. In the general case, the joint density of t_1 and t_2 given x is straightforwardly expressed as

$$f(t_1,t_2|x)=\int_0^\infty\int_0^\infty f_1(t_1|x,v_1)f_2(t_2|t_1,x,v_2)\quad dG(v_1,v_2)$$

in obvious notation. Note that if one allows for more than two consecutive spells then in practice there may be initial-conditions problems, as one may not observe the duration of the first spell.

If both durations can be followed until completion, then the data provide the joint distribution $F(t_1,t_2|x)$. Honoré (1993) shows that this model is identified from these data, under some conditions. For both marginal hazard functions in this model we make regularity assumptions corresponding to Assumptions 1–4, and we adopt regularity Assumption 10. In addition, we adopt the equivalents of Assumptions 5, 6b, and 7 on $v_i, \theta_{0,i}$, and ψ_i.[67] We also normalize the function ξ, and we replace the equivalent of Assumption 8 by a slightly different assumption,

Assumption 11 : Normalization. *For some a priori chosen* t_0*, there holds that* $\xi(t_0) = 1$.

Assumption 12 : Tails of the joint unobserved heterogeneity distribution. $\mathrm{E}(v_1) < \infty$ *and* $\mathrm{E}(v_1 v_2) < \infty$.

Sufficient for Assumption 12 is that $\mathrm{E}(v_i^2) < \infty$ for $i = 1, 2$. In sum, we adopt Assumptions 1–4, the equivalents of Assumptions 5, 6b, and 7, and Assumptions 10–12.

Here, as in the model with successive durations and $v_1 \neq v_2$ (Subsection 8.2.2), identification requires assumptions on the tails of the distributions of v_1 and v_2 (notably, finiteness of moments), and it requires that the individual hazards are proportional in t and x. It is plausible that these assumptions can be substantially weakened if the data provide multiple observations on t_1, t_2 for each v_1, v_2 pair (see Woutersen, 2000, for results).

9.2 Endogenous shocks

In this subsection we examine bivariate duration models with the property that the hazard of the duration t_2 moves to another level at the moment at which the other duration t_1 is completed, with mutually related unobserved heterogeneity terms. Recall from Section 7 that such models have been used to study the effect of punishments and training on the exit rate out of unemployment and the effect of marriage dissolution on the death rate. Also recall that the estimate of the change of the hazard is biased if one ignores the spurious dependence from related unobserved determinants. Finally, recall that we need to rule out anticipations of the realizations of t_1, but the individual is allowed to know the (determinants of the) probability distribution of t_1.

[67] In fact, the differentiability condition in Assumption 6b can be weakened to continuity here.

We adopt a framework where the two durations start at the same point of time, and where the realization of t_1 affects the shape of the hazard of t_2 from t_1 onwards. The data provide observations of t_2 and x. If t_1 is completed before t_2 then we also observe t_1; if not then we merely observe that t_1 exceeds t_2. The model and data are thus distinctly asymmetric in the two durations. Somewhat loosely, one may say that t_2 is the "main" duration, or the "endogenous duration of interest", whereas t_1 is an "explanatory" duration, and the causal effect of t_1 on t_2 is the "treatment effect".

In terms of the hazards, the model specification reads

$$\begin{aligned} \theta_1(t_1|x,v_1) &= \psi_1(t_1)\cdot\theta_{0,1}(x)\cdot v_1 \\ \theta_2(t_2|t_1,x,v_2) &= \psi_2(t_2)\cdot\theta_{0,2}(x)\cdot e^{\delta \mathrm{I}(t_1<t_2)}\cdot v_2 \end{aligned} \tag{34}$$

where I(.) denotes the indicator function, which is 1 if its argument is true and 0 otherwise. If v_1 and v_2 are independent, then, conditional on x, the durations t_1 and t_2 are only dependent if $\delta \neq 1$. In the general case, the joint density of t_1 and t_2 given x is straightforwardly derived as in the previous subsection.

In a recent working paper, Abbring and Van den Berg (2000a) provide identification results for this model. In fact, they allow δ to depend on past observables. These results are similar to those for Subsection 9.1 in that they require independence of x from v_1, v_2, and they require an assumption on the first moments of v_1, v_2. If multiple observations are available for each v_1, v_2 pair then such assumptions are not needed.

Contrary to models of binary treatments and binary outcomes, the treatment effect δ is identified without the need to rely on exclusion restrictions or parametric functional-form assumptions regarding the distribution of v_1, v_2. In particular, the set of explanatory variables affecting $\theta_{0,1}$ does not have to be larger than the set affecting $\theta_{0,2}$, and the joint distribution of v_1, v_2 can be any member of a broad nonparametric class of distributions. These results imply that the timing of events conveys useful information on the treatment effect. This information is discarded in a binary framework. In conclusion, duration analysis is useful for the study of treatment effects in non-experimental settings.[68,69]

10 Conclusions and recommendations

Since the early 1980s, the econometric analysis of duration variables has become widespread. This chapter has provided an overview of duration analysis, with an emphasis on the specification and identification of duration models, and with special attention to models for multiple durations.

We have seen that the hazard function of the duration distribution is the focal point and basic building block of econometric duration models. Properties of the duration distribution are generally discussed in terms of properties of the hazard function. The individual hazard function and the way it depends on its determinants are the "parameters of interest". This approach is dictated by economic theory. Theories that aim at explaining durations focus on the rate at which the subject leaves the state at a certain duration given that the subject has not done so yet. In particular, they explain this exit rate in terms of external conditions at the point of time corresponding to that duration and in terms of the underlying economic behavior of the subject given that he is still in the state at that duration.

The Mixed Proportional Hazard model and its special cases are by far the most popular duration models based on a specification of the hazard function. We have seen that the recent mathematical-statistical literature on counting processes has formulated precise conditions under which time-varying explanatory variables can be included in MPH models in such a way that one can still perform valid econometric inference with standard methods. Specifically, these variables have to be "predictable" stochastic processes. Here, "predictability" is a rather technical concept with a meaning similar to that of weak exogeneity.

The MPH model and its special cases are often regarded to be useful reduced-

[68] The model of this subsection does not allow the size of the treatment effect to depend on unobserved heterogeneity. Given the recent interest in heterogeneity of treatment effects (see e.g. Heckman, LaLonde and Smith, 1999), it is a challenge for future research to incorporate this into duration analysis. See Abbring and Van den Berg (2000a) for results on this.

[69] Robins (1998) analyzes treatment effects in a different type of duration models where unobserved determinants of the duration of interest may vary over time and may depend on the treatment.

form models for duration analysis. The resulting estimates are then interpreted with the help of some economic theory. Unfortunately, the proportionality assumption of the (M)PH model can in general not be justified on economic-theoretical grounds. However, if the optimal strategy of the individual is myopic (e.g. because of repeated search, or because the discount rate is infinite), then this proportionality can often be deduced from economic theory.

The MPH model is nonparametrically identified from single-spell data, given an assumption on the tail of the unobserved heterogeneity distribution, like finiteness of its mean. However, the model is nonparametrically unidentified if such an assumption is dropped. Moreover, the adoption of a model that is observationally equivalent to (but different from) the true model leads to incorrect inference on the parameters of interest. This is bad news, as it is often difficult to make any justified assumption on the tail of the unobserved heterogeneity distribution. In applications where the unobserved heterogeneity term represents an important economic variable, economic theory might provide a justification of the finite mean assumption.

Let the finite mean assumption be satisfied. The observed hazard function of the duration given the observed explanatory variables is nonproportional, meaning that it cannot be expressed as a product of a term depending only on the elapsed duration and a term depending only on the observed explanatory variables. With single-spell data, the unobserved heterogeneity distribution in MPH models is identified from the interaction between the duration and the explanatory variables in the observed hazard, or, in other words, from the observed type of nonproportionality of the observed hazard. However, unobserved heterogeneity can not generate just any type of interaction. The class of models for the observed hazard that is generated by MPH models is smaller than the general class of interaction models for the observed hazard. In other words, the MPH model is overidentified with single-spell data.

In MPH models, the sign of the interaction between the duration and the explanatory variables in the observed hazard is affected by the type of unobserved heterogeneity distribution. However, under weak conditions, the sign is always negative at small durations regardless of the type of heterogeneity distribution. If unobserved heterogeneity has a Gamma distribution, then the interaction is negative at all durations and all values of the systematic part of the hazard function. If unobserved heterogeneity has a discrete distribution with two positive mass points then the interaction is negative at small durations and positive at large durations.

In MPH models, the effect of an explanatory variable on the observed hazard can be negative at some durations even if the explanatory variable has a positive effect on the underlying individual (or systematic) hazard. This means that it is not possible to deduce the sign of the effect of the explanatory variable on the underlying individual hazard from the observed effect of the variable on the observed hazard at certain durations. Fortunately, this remarkable effect can only occur for some local duration intervals.

By now, there is overwhelming evidence that with single-spell data, minor changes in the assumed parametric specification of the MPH model, while leading to a similar over-all fit, may produce very different parameter estimates. Also, very different models may generate similar data. Estimation results from single-spell data are sensitive to misspecification of the functional forms associated with the model determinants, and this sensitivity is stronger than usual in econometrics. In the absence of strong prior information on the model determinants, single-spell data do not enable a robust assessment of the relative importance of these determinants as explanations of random variation in the observed durations. Therefore, interpretations based on estimation results are often unstable and should be performed with extreme caution.

In biostatistics, this state of affairs has led to a renewed interest in Accelerated Failure Time models as alternative reduced-form duration models for the analysis of single-spell duration data. From an econometric point of view, the AFT approach is unsatisfactory, because it does not focus on the parameters of the individual hazard as the parameters of interest. However, if one is only interested in the sign or significance of a covariate effect on the individual durations then the AFT approach may be useful.

In practice, it may be useful to exploit predictions from the underlying economic theory when specifying the duration model, by imposing these as restrictions on the functional form of the heterogeneity distribution or the baseline hazard. It may be even more useful to look for data with multiple spells (see below). Now suppose that these options are not available. Concerning the baseline hazard, the conceived wisdom is that a piecewise constant specification is then the most useful. Such a specification is flexible and convenient from a computational point of view. Concerning the unobserved heterogeneity distribution, it may be useful to start off with an informal examination of the sign of the interaction in the observed hazard. If it is negative at all durations then a Gamma distribution may give a better fit whereas if it is positive at large durations then a discrete distribution may give a better fit.

By now, the empirical analysis of MPH models with multi-spell duration data is widespread. Basically, if two observations are available for each unobserved heterogeneity value, then the identification of the model does not require an untestable assumption on the tail of the unobserved heterogeneity distribution anymore, and, perhaps even more importantly, observed and unobserved explanatory variables are allowed to be dependent. The identification of this distribution does not come anymore from the interaction between the duration and the observable explanatory variables in the observed hazard. Data on multiple spells for the same individual therefore remove the identification problems associated with single-spell data. Moreover, a consensus has emerged that multi-spell data allow for reliable inference that is robust with respect to the specification of the unobserved heterogeneity distribution. Multi-spell duration data make duration analysis more similar to dynamic panel data analysis. It should however be stressed that the analysis of multi-spell data is particularly sensitive to censoring.

The chapter pays special attention to models for multiple durations. Here, the marginal duration distributions need not be the same. In general one may think of many different ways to model a relation between duration variables. In the applied econometric literature on the estimation of multiple-duration models, the range of different models is actually not so large. Typically, the models allow for dependence between the duration variables by way of their unobserved determinants, with each single duration following its own MPH model. In addition to this, the model may allow for a "causal" effect of one duration on the other, as motivated by an underlying economic theory. The first popular type of causal effect concerns an effect of a realized past duration on the current hazard. Basically, this is modeled by including the realized past duration as an additional covariate in the hazard for the current duration. The second popular type of causal effect concerns situations where two durations occur simultaneously, and where the realization of one duration variable has an immediate effect on the hazard of the other duration variable. This includes models of treatment effects in the presence of selectivity and in the absence of exclusion restrictions.

For such models, identification results have been derived which are similar in contents to those for MPH models with single-spell data. The identification conditions can be weakened substantially if multiple observations are available for each value of the heterogeneity pair, or if cross-restrictions are imposed on the distributions of the two durations in the multiple duration model.

The multiple-duration model where the marginal duration distributions each satisfy an MPH specification, and the durations can only be dependent by way of their unobserved determinants, is called the Multivariate Mixed Proportional Hazard (MMPH) model. In the empirical analysis with such models it is important to assume a genuine multivariate distribution for the unobserved heterogeneity terms. Here, "genuine" means that there is no deterministic relation between any two heterogeneity terms. More restrictive specifications, like the one-factor loading specification, impose cross-restrictions on the marginal duration distributions and the dependence of the durations. In such cases, if the data provide evidence for unobserved heterogeneity in the marginal duration distributions, then the model implies that these durations must be dependent. Similarly, in such cases, if the durations are independent, then the model implies that there is no unobserved heterogeneity for at least one of the durations.

Factor loading specifications have been popular because they restrict the number of unknown parameters, leading to a sparse specification, and they limit the computational burden of the estimation of the model. However, the latter can also be achieved by adopting a (multidimensional) discrete distribution for the unobserved heterogeneity terms. In fact, discrete heterogeneity distributions are particularly flexible, in the sense that they are able to generate a relatively wide range of values for the association measures of the corresponding durations. In empirical applications with MMPH models, it is therefore useful for computational reasons and for reasons of flexibility to assume a multidimensional discrete distribution for the unobserved heterogeneity terms. One may then try to increase the number of mass points. If the number of duration types is relatively large then one may reduce the number of parameters of the multidimensional discrete distribution somewhat by imposing, say, a two-factor loading structure.

References

Abbring, J.H. (1998), "Treatment effects in duration models and Granger non-causality", Working paper, Free University, Amsterdam.

Abbring, J.H. and G.J. van den Berg (1998), "Limit results for unobserved heterogeneity distributions in mixed proportional hazard models", Working paper, Free University Amsterdam, Amsterdam.

Abbring, J.H. and G.J. van den Berg (2000a), "The nonparametric identification of treatment effects in duration models", Working paper, Free University Amsterdam, Amsterdam.

Abbring, J.H. and G.J. van den Berg (2000b), "The non-parametric identification of the mixed-proportional-hazard competing-risks model", Working paper, Free University Amsterdam, Amsterdam.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (1997), "The effect of unemployment insurance sanctions on the transition rate from unemployment to employment", Working paper, Tinbergen Institute, Amsterdam.

Andersen, P.K., M.W. Bentzon, and J.P. Klein (1996), "Estimating the survival function in the proportional hazards regression model: a study of the small sample size properties", *Scandinavian Journal of Statistics*, 23, 1–12.

Andersen, P.K. and Ø. Borgan (1985), "Counting process models for life history data: a review (with discussion)", *Scandinavian Journal of Statistics*, 12, 97–158.

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1993), *Statistical Models Based on Counting Processes*, Springer, New York.

Anderson, J.E., T.A. Louis, N.V. Holm, and B. Harvald (1992), "Time-dependent association measures for bivariate survival distributions", *Journal of the American Statistical Association*, 87, 641–650.

Anti Nilsen, Ø. and F. Schiantarelli (1998), "Zeroes and lumps in investment: empirical evidence on irreversibilities and non-convexities", Working paper, University of Bergen, Bergen.

Antonides, G. (1988), *Scrapping a durable consumption good*, Erasmus Universiteit Rotterdam, Rotterdam.

Arjas, E. (1989), "Survival models and martingale dynamics (with discussion)", *Scandinavian Journal of Statistics*, 16, 177–225.

Arroyo, C.R. and J. Zhang (1997), "Dynamic microeconomic models of fertility choice: a survey", *Journal of Population Economics*, 10, 23–65.

Baker, M. and A. Melino (2000), "Duration dependence and nonparametric heterogeneity: a Monte Carlo study", *Journal of Econometrics*, forthcoming.

Bauwens, L. and P. Giot (1998), "The logarithmic ACD model: an application to the bid/ask quote process of two NYSE stocks", Working paper, CORE, Louvain-la-Neuve.

Blanchard, O.J. and P. Diamond (1994), "Ranking, unemployment duration, and wages", *Review of Economic Studies*, 61, 417–434.

Boizot, C., J.M. Robin, and M. Visser (1997), "The demand for food products: An analysis of interpurchase times and purchased quantities", Working paper, CREST, Paris.

Bonnal, L., D. Fougère, and A. Sérandon (1997), "Evaluating the impact of French employment policies on individual labour market histories", *Review of Economic Studies*, 64, 683–713.

Bontemps, C., J.M. Robin, and G.J. van den Berg (1999), "An empirical equilibrium job search model with search on the job and heterogeneous workers and firms", *International Economic Review*, 40, 1039–1074.

Bontemps, C., J.M. Robin, and G.J. van den Berg (2000), "Equilibrium search with continuous productivity dispersion: theory and non-parametric estimation", *International Economic Review*, 41, 305–358.

Bowlus, A.J., N.M. Kiefer, and G.R. Neumann (2001), "Equilibrium search models and the transition from school to work", *International Economic Review*, 42, forthcoming.

Bretagnolle, J. and C. Huber-Carol (1988), "Effect of omitting covariates in Cox's model for survival data", *Scandinavian Journal of Statistics*, 15, 125–138.

Burgess, S. (1989), "The estimation of structural models of unemployment duration with on-the-job search", Working paper, University of Bristol, Bristol.

Butler, J.S., K.H. Anderson, and R.V. Burkhauser (1986), "Testing the relationship between work and health", *Economics Letters*, 20, 383–386.

Butler, J.S., K.H. Anderson, and R.V. Burkhauser (1989), "Work and health after retirement", *Review of Economics and Statistics*, 71, 46–53.

Card, D. and D. Sullivan (1988), "Measuring the effect of subsidized training programs on movements in and out of employment", *Econometrica*, 56, 497–530.

Carling, K. and T. Jacobson (1995), "Modeling unemployment duration in a dependent competing risks framework: identification and estimation", *Lifetime Data Analysis*, 1.

Chamberlain, G. (1985), "Heterogeneity, omitted variable bias, and duration dependence", in J.J. Heckman and B. Singer, editors, *Longitudinal analysis of labor market data*, Cambridge University Press, Cambridge.

Chesher, A. and T. Lancaster (1983), "The estimation of models of labour market behaviour", *Review of Economic Studies*, 50, 609–624.

Coleman, T.S. (1990), "Unemployment behaviour: evidence from the CPS work experience survey", in Y. Weiss and G. Fishelson, editors, *Advances in the theory and measurement of unemployment*, Macmillan, Basingstoke.

Cox, D.R. (1972), "Regression models and life tables (with discussion)", *Journal of the Royal Statistical Society Series B*, 34, 187–220.

Dabrowska, D.M. (1987), "Non-parametric regression with censored survival time data", *Scandinavian Journal of Statistics*, 14, 181–197.

Devine, T.J. and N.M. Kiefer (1991), *Empirical Labor Economics*, Oxford University Press, Oxford.

Diebold, F.X. and G.D. Rudebusch (1990), "A nonparametric investigation of duration dependence in the American business cycle", *Journal of Political Economy*, 98, 596–616.

Eberwein, C., J.C. Ham, and R.J. LaLonde (1997), "The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: evidence from experimental data", *Review of Economic Studies*, 64, 655–682.

Elbers, C. and G. Ridder (1982), “True and spurious duration dependence: The identifiability of the proportional hazard model”, *Review of Economic Studies*, 49, 403–410.

Engberg, J. (1991), “The impact of unemployment benefits on job search: structural unobserved heterogeneity and spurious spikes”, Working paper, Carnegie-Mellon University, Pittsburgh.

Engberg, J., P. Gottschalk, and D. Wolf (1990), “A random-effects logit model of work-welfare transitions”, *Journal of Econometrics*, 43, 63–75.

Engle, R. and J. Russell (1998), “Autoregressive conditional duration; a new model for irregularly spaced transaction data”, *Econometrica*, 66, 1127–1162.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications II*, Wiley, New York.

Fleming, T.R. and D.P. Harrington (1991), *Counting Processes and Survival Analysis*, Wiley, New York.

Flinn, C.J. (1996), “Labor market structure and welfare: a comparison of Italy and the U.S.”, Working paper, New York University, New York.

Flinn, C.J. and J.J. Heckman (1982a), “New methods for analyzing structural models of labor force dynamics”, *Journal of Econometrics*, 18, 115–168.

Flinn, C.J. and J.J. Heckman (1982b), “Models for the analysis of labor force dynamics”, in R. Basmann and G. Rhodes, editors, *Advances in Econometrics, volume 1*, JAI Press, Greenwich.

Flinn, C.J. and J.J. Heckman (1983), “Are unemployment and out of the labor force behaviorally distinct labor force states?”, *Journal of Labor Economics*, 1, 28–42.

Florens, J.P. and D. Fougère (1996), “Noncausality in continuous time”, *Econometrica*, 64, 1195–1212.

Freund, J.E. (1961), “A bivariate extension of the exponential distribution”, *Journal of the American Statistical Association*, 56, 971–977.

Gail, M.H., S. Wieand, and S. Piantadosi (1984), “Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates”, *Biometrika*, 71, 431–444.

Garcia-Perez, J.I. (1998), "Non-stationary job search with firing: a structural estimation", Working paper, CEMFI, Madrid.

Genest, C. and J. MacKay (1986), "The joy of copulas: bivariate distributions with uniform marginals", *The American Statistician*, 40, 280–283.

Gönül, F. and K. Srinivasan (1993), "Consumer purchase behavior in a frequently bought product category: estimation issues and managerial insights from a hazard function model with heterogeneity", *Journal of the American Statistical Association*, 88, 1219–1227.

Gritz, R.M. (1993), "The impact of training on the frequency and duration of employment", *Journal of Econometrics*, 57, 21–51.

Guo, G. and G. Rodríguez (1992), "Estimating a multivariate proportional hazard model for clustered data using the EM algorithm, with an application to child survival in Guatemala", *Journal of the American Statistical Association*, 87, 969–976.

Hahn, J. (1994), "The efficiency bound of the mixed proportional hazard model", *Review of Economic Studies*, 61, 607–629.

Ham, J.C. and R.J. LaLonde (1996), "The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training", *Econometrica*, 64, 175–205.

Ham, J.C. and S.A. Rea (1987), "Unemployment insurance and male unemployment duration in Canada", *Journal of Labor Economics*, 5, 325–353.

Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47, 153–161.

Heckman, J.J. (1991), "Identifying the hand of the past: distinguishing state dependence from heterogeneity", *American Economic Review*, 81 (supplement), 71–79.

Heckman, J.J. and G.J. Borjas (1980), "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence", *Economica*, 47, 247–283.

Heckman, J.J. and B.E. Honoré (1989), "The identifiability of the competing risks model", *Biometrika*, 76, 325–330.

Heckman, J.J., V.J. Hotz, and J.R. Walker (1985), "New evidence on the timing and spacing of births", *American Economic Review*, 75, 179–184.

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999), "The economics and econometrics of active labor market programs", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Heckman, J.J. and B. Singer (1984a), "Econometric duration analysis", *Journal of Econometrics*, 24, 63–132.

Heckman, J.J. and B. Singer (1984b), "The identifiability of the proportional hazard model", *Review of Economic Studies*, 51, 231–241.

Heckman, J.J. and B. Singer (1984c), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", *Econometrica*, 52, 271–320.

Heckman, J.J. and C.R. Taber (1994), "Econometric mixture models and more general models for unobservables in duration analysis", *Statistical Methods in Medical Research*, 3, 279–302.

Heckman, J.J. and J.R. Walker (1987), "Using goodness of fit and other criteria to choose among competing duration models: a case study of the Hutterite data", in C. Clogg, editor, *Sociological Methodology 1987*, American Sociological Association, Washington.

Heckman, J.J. and J.R. Walker (1990), "The relationship between wages and income and the timing and spacing of births", *Econometrica*, 58, 1411–1441.

Honoré, B.E. (1991), "Identification results for duration models with multiple spells or time-varying covariates", Working paper, Northwestern University, Evanston.

Honoré, B.E. (1993), "Identification results for duration models with multiple spells", *Review of Economic Studies*, 60, 241–246.

Horowitz, J.L. (1996), "Semiparametric estimation of a regression model with an unknown transformation of the dependent variable", *Econometrica*, 64, 103–137.

Horowitz, J.L. (1999), "Semiparametric estimation of a proportional hazard model with unobserved heterogeneity", *Econometrica*, 67, 1001–1028.

Horvath, W.J. (1968), "A statistical model for the duration of wars and strikes", *Behavioral Science*, 13, 24.

Hougaard, P. (1986), "A class of multivariate failure time distributions", *Biometrika*, 73, 671–678.

Hougaard, P. (1987), "Modelling multivariate survival", *Scandinavian Journal of Statistics*, 14, 291–304.

Hougaard, P. (1991), "Modelling heterogeneity in survival data", *Journal of Applied Probability*, 28, 695–701.

Hougaard, P., B. Harvald, and N.V. Holm (1992a), "Measuring the similarities between the lifetimes of adult Danish twins born between 1881-1930", *Journal of the American Statistical Association*, 87, 17–24.

Hougaard, P., B. Harvald, and N.V. Holm (1992b), "Assessment of dependence in the life times of twins", in J.P. Klein and P.K. Goel, editors, *Survival Analysis: State of the Art*, Kluwer Academic Publishers, Dordrecht.

Hougaard, P., P. Myglegaard, and K. Borch-Johnsen (1994), "Heterogeneity models of disease susceptibility, with an application to diabetic nephropathy", *Biometrics*, 50, 1178–1188.

Jovanovic, B. (1984), "Wages and turnover: a parametrization of the job-matching model", in G.R. Neumann and N. Westergård-Nielsen, editors, *Studies in Labor Market Dynamics*, Springer-Verlag, Heidelberg.

Kalbfleisch, J.D. and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Keiding, N. (1998), "Selection effects and nonproportional hazards in survival models and models for repeated events", Working paper, University of Copenhagen, Copenhagen.

Keiding, N., P.K. Andersen, and J.P. Klein (1997), "The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates", *Statistics in Medicine*, 16, 215–224.

Kendall, M.G. (1962), *Rank correlation methods*, Griffin, London.

Kennan, J.F. (1985), "The duration of contract strikes in U.S. manufacturing", *Journal of Econometrics*, 28, 5–28.

Kiefer, N.M. (1988), "Economic duration data and hazard functions", *Journal of Economic Literature*, 26, 646–679.

Klaassen, C.A.J. and A.J. Lenstra (1998), "The information for the treatment effect in the mixed proportional hazards model vanishes", Working paper, University of Amsterdam, Amsterdam.

Koning, P., G.J. van den Berg, G. Ridder, and K. Albæk (2000), "The relation between wages and labor market frictions: An empirical analysis based on matched worker-firm data", in H. Bunzel et al., editors, *Panel data and structural labour market models*, North Holland, Amsterdam.

Kortram, R.A., A.J. Lenstra, G. Ridder, and A.C.M. van Rooij (1995), "Constructive identification of the mixed proportional hazards model", *Statistica Neerlandica*, 49, 269–281.

Lancaster, T. (1979), "Econometric methods for the duration of unemployment", *Econometrica*, 47, 939–956.

Lancaster, T. (1983), "Generalized residuals and heterogeneous duration models: the exponential case", *Bulletin of Economic Research*, 35, 71–86.

Lancaster, T. (1985a), "Simultaneous equations models in applied search theory", *Journal of Econometrics*, 28, 155–169.

Lancaster, T. (1985b), "Generalised residuals and heterogeneous duration models – with applications to the Weibull model", *Journal of Econometrics*, 28, 113–126.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Lancaster, T. (2000a), "The incidental parameter problem since 1948", *Journal of Econometrics*, 95, 391–413.

Lancaster, T. (2000b), "Some econometrics of scarring", in *Nonlinear Statistical Inference: Essays in Honor of T. Amemiya*, Cambridge University Press, Cambridge.

Lenstra, A.J. and A.C.M. van Rooij (1998), "Nonparametric estimation of the mixed proportional hazards model", Working paper, Free University, Amsterdam.

Li, Y., J.P. Klein, and M.L. Moeschberger (1993), "Effects of model misspecification in estimating covariate effects in survival analysis for small sample sizes", Working paper, Medical College of Wisconsin, Milwaukee.

Lillard, L.A. (1993), "Simultaneous equations for hazards", *Journal of Econometrics*, 56, 189–217.

Lillard, L.A. and C.W.A. Panis (1996), "Marital status and mortality: the role of health", *Demography*, 33, 313–327.

Lillard, L.A. and C.W.A. Panis (1998), "Panel attrition from the PSID", *Journal of Human Resources*, 33, 437–457.

Lindeboom, M. and M. Kerkhofs (2000), "Time patterns of work and sickness absence: unobserved workplace effects in a multi-state duration model", *Review of Economics and Statistics*, forthcoming.

Lindeboom, M. and G.J. van den Berg (1994), "Heterogeneity in bivariate duration models: the importance of the mixing distribution", *Journal of the Royal Statistical Society Series B*, 56, 49–60.

Lindstrom, D.P. (1996), "Economic opportunity in Mexico and return migration from the United States", *Demography*, 33, 357–374.

McCall, B.P. (1996), "The identifiability of the mixed proportional hazards model with time-varying coefficients", *Econometric Theory*, 12, 733–738.

Meghir, C. and E. Whitehouse (1997), "Labour market transitions and retirement of men in the UK", *Journal of Econometrics*, 79, 327–354.

Melino, A. and G.T. Sueyoshi (1990), "A simple approach to the identifiability of the proportional hazards model", *Economics Letters*, 33, 63–68.

Meyer, B.D. (1995), "Semiparametric estimation of hazard models", Working paper, Northwestern University, Evanston.

Mortensen, D.T. (1986), "Job search and labor market analysis", in O. Ashenfelter and R. Layard, editors, *Handbook of Labor Economics*, North-Holland, Amsterdam.

Mortensen, D.T. and C.A. Pissarides (1999), "New developments in models of search in the labor market", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Moscarini, G. (1997), "Unobserved heterogeneity and unemployment duration: a fallacy of composition", Working paper, Yale University, New Haven.

Mroz, T. and D. Weir (1990), "Structural change in life cycle fertility during the fertility transition: France before and after the revolution of 1789", *Population Studies*, 44, 61–87.

Narendranathan, W. (1993), "Job search in a dynamic environment - an empirical analysis", *Oxford Economic Papers*, 45, 1–22.

Narendranathan, W. and S.J. Nickell (1985), "Modelling the process of job search", *Journal of Econometrics*, 28, 28–49.

Neumann, G.R. (1997), "Search models and duration data", in M.H. Pesaran, editor, *Handbook of Applied Econometrics: Microeconometrics*, Basil Blackwell, Oxford.

Newman, J.L. and C.E. McCullogh (1984), "A hazard rate approach to the timing of births", *Econometrica*, 52, 939–961.

Ng, E.T.M. and R.J. Cook (1997), "Modeling two-state disease processes with random effects", *Lifetime Data Analysis*, 3, 315–335.

Nickell, S.J. (1979), "Estimating the probability of leaving unemployment", *Econometrica*, 47, 1249–1266.

Nielsen, G.G., R.D. Gill, P.K. Andersen, and T.I.A. Sørensen (1992), "A counting process approach to maximum likelihood estimation in frailty models", *Scandinavian Journal of Statistics*, 19, 25–43.

Oakes, D. (1989), "Bivariate survival models induced by frailties", *Journal of the American Statistical Association*, 84, 487–493.

Omori, Y. (1997), "Stigma effects of nonemployment", *Economic Inquiry*, 35, 394–416.

Pakes, A. and M. Schankerman (1984), "The rate of obsolescence of knowledge, research gestation lags, and the private rate of return to research resources", in Z. Griliches, editor, *Patents, R& D and Productivity*, University of Chicago Press, Chicago.

Pelz, C.J. and J.P. Klein (1996), "Analysis of survival data: a comparison of three major statistical packages (SAS, SPSS, BMDP)", Working paper, Medical

College of Wisconsin, Milwaukee.

Petersen, J.H. (1996), "A litter frailty model", Working paper, University of Copenhagen, Copenhagen.

Pissarides, C.A. (1990), *Equilibrium Unemployment Theory*, Basil Blackwell, Oxford.

Ridder, G. (1984), "The distribution of single-spell duration data", in G.R. Neumann and N. Westergård-Nielsen, editors, *Studies in Labor Market Dynamics*, Springer-Verlag, Heidelberg.

Ridder, G. (1987), "The sensitivity of duration models to misspecified unobserved heterogeneity and duration dependence", Working paper, Groningen University, Groningen.

Ridder, G. (1990), "The non-parametric identification of generalized accelerated failure-time models", *Review of Economic Studies*, 57, 167–182.

Ridder, G. and I. Tunalı (1999), "Stratified partial likelihood estimation", *Journal of Econometrics*, forthcoming.

Ridder, G. and G.J. van den Berg (1997), "Empirical equilibrium search models", in D.M. Kreps and K.F. Wallis, editors, *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, Cambridge.

Ridder, G. and G.J. van den Berg (1998), "Estimating measures of labor market imperfection for five OECD countries, using aggregate data in an equilibrium search framework", Working paper, Free University Amsterdam, Amsterdam.

Robins, J.M. (1998), "Structural nested failure time models", in P. Armitage and T. Colton, editors, *The Encyclopedia of Biostatistics*, John Wiley and Sons, Chichester.

Robins, J.M. and S. Greenland (1989), "The probability of causation under a stochastic model for individual risk", *Biometrics*, 45, 1125–1138.

Rosholm, M. (1997), "The risk of marginalization in the labour market: application of a three state dependent competing risks duration model", Working paper, University of Aarhus, Aarhus.

Rust, J. (1994), "Structural estimation of Markov decision processes", in R.F. Engle and D.L. McFadden, editors, *Handbook of Econometrics, Volume IV*, North-Holland, Amsterdam.

Ryu, K. (1993), “Structural duration analysis of management data”, *Journal of Econometrics*, 57, 91–115.

Sastry, N. (1997), “A nested frailty model for survival data, with an application to the study of child survival in Northeast Brazil”, *Journal of the American Statistical Association*, 92, 426–435.

Snyder, D.L. and M.I. Miller (1991), *Random Point Processes in Time and Space*, Springer-Verlag, Heidelberg.

Solomon, P.J. (1984), “Effect of misspecification of regression models in the analysis of survival data”, *Biometrika*, 71, 291–298 (Correction, 1986, page 245).

Solon, G. (1985), “Work incentive effects of taxing unemployment benefits”, *Econometrica*, 53, 295–306.

Thomas, J.M. (1998), “The role of selective job search in UK unemployment”, *Economic Journal*, 108, 646–664.

Trussell, J. and T. Richards (1985), “Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure”, in N. Tuma, editor, *Sociological Methodology 1985*, Jossey-Bass, San Francisco.

Van den Berg, G.J. (1990a), “Nonstationarity in job search theory”, *Review of Economic Studies*, 57, 255–277.

Van den Berg, G.J. (1990b), “Search behaviour, transitions to nonparticipation and the duration of unemployment”, *Economic Journal*, 100, 842–865.

Van den Berg, G.J. (1990c), “The effect of an increase of the rate of arrival of job offers on the duration of unemployment”, Working paper, Groningen University, Groningen.

Van den Berg, G.J. (1992), “Nonparametric tests for unobserved heterogeneity in duration models”, Working paper, Free University, Amsterdam.

Van den Berg, G.J. (1995), “Explicit expressions for the reservation wage path and the unemployment duration density in nonstationary job search models”, *Labour Economics*, 2, 187–198.

Van den Berg, G.J. (1997), “Association measures for durations in bivariate hazard rate models”, *Journal of Econometrics*, 79, 221–245.

Van den Berg, G.J. (1999), "Empirical inference with equilibrium search models of the labor market", *Economic Journal*, 109, F283–F306.

Van den Berg, G.J., A. Holm, and J.C. van Ours (1999), "Do stepping-stone jobs exist? early career paths in the medical profession", Working paper, Tinbergen Institute, Amsterdam.

Van den Berg, G.J. and M. Lindeboom (1998), "Attrition in panel survey data and the estimation of multi-state labor market models", *Journal of Human Resources*, 33, 458–478.

Van den Berg, G.J., M. Lindeboom, and G. Ridder (1994), "Attrition in longitudinal panel data, and the empirical analysis of dynamic labour market behaviour", *Journal of Applied Econometrics*, 9, 421–435.

Van den Berg, G.J. and G. Ridder (1998), "An empirical equilibrium search model of the labor market", *Econometrica*, 66, 1183–1221.

Van den Berg, G.J., B. van der Klaauw, and J.C. van Ours (1998), "Punitive sanctions and the transition rate from welfare to work", Working paper, Tinbergen Institute, Amsterdam.

Van den Berg, G.J. and J.C. van Ours (1996), "Unemployment dynamics and duration dependence", *Journal of Labor Economics*, 14, 100–125.

Van der Klaauw, W. (1996), "Female labour supply and marital status decisions: a life-cycle model", *Review of Economic Studies*, 63, 199–235.

Vaupel, J.W., K.G. Manton, and E. Stallard (1979), "The impact of heterogeneity in individual frailty on the dynamics of mortality", *Demography*, 16, 439–454.

Vilcassim, N.J. and D.C. Jain (1991), "Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity", *Journal of Marketing Research*, 28, 29–41.

Visser, M. (1996), "Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS", *Biometrika*, 83, 507–518.

Wang, S.T., J.P. Klein, and M.L. Moeschberger (1995), "Semi-parametric estimation of covariate effects using the Positive Stable frailty model", *Applied Stochastic Models and Data Analysis*, 11, 121–133.

Wolpin, K.I. (1987), "Estimating a structural job search model: the transition from school to work", *Econometrica*, 55, 801–818.

Wolpin, K.I. (1995), *Empirical methods for the study of labor force dynamics*, Harwood Academic Publishers, Luxembourg.

Woutersen, T.M. (2000), "Consistent estimators for panel duration data with endogenous censoring and endogenous regressors", Working paper, Brown University, Providence.

Xue, X. and R. Brookmeyer (1996), "Bivariate frailty model for the analysis of multivariate survival time", *Lifetime Data Analysis*, 2, 277–289.

Yamaguchi, K. (1986), "Alternative approaches to unobserved heterogeneity in the analysis of repeatable events", in N.B. Tuma, editor, *Sociological Methodology 1986*, Jossey-Bass, Washington D.C.

Yamaguchi, K. (1991), *Event History Analysis*, Sage, Newbury Park.

Yashin, A.I. and I.A. Iachine (1997), "How frailty models can be used in evaluating longevity limits", *Demography*, 34, 31–48.

Yashin, A.I. and I.A. Iachine (1999), "What difference does the dependence between durations make? Insights for population studies of aging", *Lifetime Data Analysis*, 5, 5–22.

Yoon, B.J. (1981), "A model of unemployment duration with variable search intensity", *Review of Economics and Statistics*, 63, 599–609.

Yoon, B.J. (1985), "A non-stationary hazard function of leaving unemployment for employment", *Economics Letters*, 17, 171–175.

Zahl, P.H. (1997), "Frailty modelling for the excess hazard", *Statistics in Medicine*, 16, 1573–1585.