TRANSITION MODELS IN A NON-STATIONARY ENVIRONMENT

G. W. Imbens*

Abstract—An alternative form of the proportional hazard model is proposed. It allows one to introduce correlation between exit rates at the same (calendar) time for different individuals. One can, in the context of this model, still allow for, and estimate, duration effects. These should be parametrized. These modifications to the original Cox model are possible by reversing the roles of duration and calendar time. It is argued that flexibility with respect to the effects of these macro processes is of particular relevance in economic models. An example using Dutch data on labor market transitions illustrates the idea that to ignore calendar time effects may have severe consequences for the estimation of duration dependence.

I. Introduction

THE econometric analysis of transition data has been the subject of a large and growing literature since the late seventies. Many of the theoretical concepts in this field, however, are borrowed from the biostatistical literature. This phenomenon is not restricted to the terminology (hazard and survivor functions are the most obvious examples), but extends to the models employed. One such model is the proportional hazard specification, proposed by Cox (1972, 1975). It allows one to study the effects of regressors on transition rates without specifying the form of the duration dependence. Econometrics is not biostatistics, however, and the analysis of economic issues brings with it special problems that are not necessarily satisfactorily dealt with by these models. An important econometric innovation was the introduction of unobserved heterogeneity in these models by Lancaster (1979). (See also Lancaster (1990) and Heckman and Singer (1986).) In economics it is often more difficult to control for individual differences than in a controlled hospital environment, where the data for biomedical studies are often obtained.

Received for publication April 1, 1992. Revision accepted for publication May 24, 1993.
* Harvard University.
I wish to acknowledge helpful suggests from Tony Lancaster, Geert Ridder, Larry Katz, Bob Miller, Marcel Kerkhofs, Michael Visser, an anonymous referee and the participants of seminars at CREST/ENSAE, Tilburg University and Harvard/MIT. An earlier version of this paper appeared as a CentER discussion paper under the title "Duration Models with Time-Varying Coefficients." A substantial part of the work on this paper was done while the author was at Tilburg University.

In this paper we want to point out a second implication of the lack of control in economic environments. Not only is it plausible that the population is heterogeneous, but it is also likely that the environment in which the population exists changes over time. In the analysis of panel data this has led researchers to introduce time dummies as well as individual effects. In the analysis of continuous time duration data there is no natural time period, and time dummies would therefore not constitute an appealing solution. We will derive a continuous time analogue where instead of a finite dimensional vector of nuisance parameters, we have an infinite dimensional nuisance parameter. This semi-parametric technique is obtained by reversing the role of duration and calendar time in the Cox regression model. In addition to the flexibility that the semi-parametric specification provides, this estimator can be considerably less demanding computationally than fully parametric procedures. This stems from the fact that fully parametric procedures require evaluation of the integral of the hazard rate, while the Cox regression model only requires evaluation of the hazard rate itself, which can be significantly easier if the model specification itself is in terms of the hazard function.

The key assumption is that the common calendar time effects enter the hazard function multiplicatively. This is of course a restrictive assumption, in the same way the assumption of any proportional duration effect is restrictive in the original form of the Cox regression model. However, one can still allow for time effects that vary across the population, as long as one is willing to parametrize this variation.

We also discuss how the common time-effects can be estimated in a second stage after the parameter vector has been estimated. These time effects might be of independent interest if one is interested in decomposing variation in the unemployment rate over time into variation in the inflow and variation in the outflow.

The outline of the paper is as follows: in the next section the main ideas are first discussed in the simple framework with single spells and time-invariant regressors.[1] This basic model is general enough to convey the main theoretical ideas of the paper. In the third section the analysis of the preceding sections is applied to a Dutch dataset on labor market histories. The results suggest that variation of the average rather than the conditional hazard explains most of the variation in the length of employment and unemployment spells. The final section contains the conclusion and summary of the main findings.

II. Calendar Time versus Duration Effects

Consider a world in which individuals experience two events in a particular order. The first might be labelled *entry* and the second *exit*. Examples of such events in economics are the beginning and the end of employment or unemployment spells and births and deaths of companies. At times we will come back to these examples to motivate particular modeling decisions, while at other times it might be more expedient to refer to the more abstract version of the model. We are interested in the timing of the second event, t^1, given the date that the first event occurred, t^0, and given some time-invariant characteristics of the individual which are denoted by a vector x. The duration of the spell between the first and second event will be analyzed primarily in terms of the *hazard function*, *or intensity process*, defined as:

$$\lambda(t, t^0, x) = \lim_{\delta \downarrow 0} \mathscr{P}\left[t^1 \in [t, t+\delta) | t^1 \geq t, t^0, x\right] \Big/ \delta \quad \text{for } t > t^0. \qquad (1)$$

It is a function of calendar time t, entry time t^0 and characteristics x. Often it has been assumed that the hazard function is a function of duration $t - t^0$ and characteristics x alone. A good example of this type of analysis is the paper by Lancaster (1979) that sparked off a whole literature on duration models in econometrics.

One way to analyze models of this type given a dataset[2] $(t_n^1, t_n^0, x_n)_{n=1}^N$ is by specifying a parametric form for the hazard and estimating the parameters by maximum likelihood techniques. Suppose the hazard function has the following form:

$$\lambda(t, t^0, x) = \omega(t, t^0, x; \theta) \qquad (2)$$

with ω a known function, and θ an unknown parameter. The likelihood function would, in that case, be:

$$\mathscr{L}(\theta) = \prod_{n=1}^{N} \omega(t_n^1, t_n^0, x_n; \theta) \times \exp\left[-\int_{t_n^0}^{t_n^1} \omega(s, t_n^0, x_n; \theta)\, ds\right]. \qquad (3)$$

Under standard regularity conditions the maximum likelihood estimator is consistent and asymptotically normal. Ridder (1987), among others, follows this approach. He ignores duration dependence and allows for a flexible calendar time dependence by introducing dummies for two-year intervals. The disadvantage of the model is that the functional form of the hazard function has to be specified completely. Using dummies can to some extent overcome this problem but there is no natural time period in these continuous time models, unlike in panel data analysis. In the remainder of section II we will study ways in which we can incorporate calendar time dependence in a more satisfactory way. First we look at a technique developed by Cox to allow for a very general form of duration dependence. This technique will then be adapted to make it flexible with respect to calendar time dependence.

A. *The Cox Regression Model*

To overcome the heavy reliance on knowledge of the functional form that characterized the hazard function in (2), Cox (1972, 1975) proposed the *proportional hazard model*, also known as the *Cox regression model*. In his analysis the hazard depends only on duration $t - t^0$ and characteristics x. Alternatively, one can interpret this as the special case of (1) where all t_n^0 are identical. Cox makes the assumption that the hazard rate can be factorized into a function of duration alone and a function of characteristics alone:

$$\lambda(t, t^0, x) = \lambda_0(t - t^0) \cdot \omega(x; \theta) \qquad (4)$$

[1] The assumption of time-invariant regressors is not essential. The entire analysis can be done allowing for time varying regressors, as long as they are predictable. In order to simplify notation we do not introduce this complication.

[2] In the discussion in this section the complication of *censored* observations will be ignored for expository reasons. However, the formal results towards the end of the section do allow for censoring.

The first factor, λ_0, the *baseline hazard*, is unknown to the researcher, but the second factor, $\omega(\cdot)$, is known up to a finite dimensional parameter θ. This parameter can be estimated using Cox's *partial likelihood* approach. This procedure will be discussed here in some detail as it provides insights in the way we can avoid specification of part of the model as long as other parts are parametrically specified. Later we will use the same procedure with different parts of the model unspecified. For a fuller discussion of the principle of partial likelihood see Cox and Oakes (1984) or Lancaster (1990). The basic idea is that we *condition* on the durations which are mainly determined by the baseline hazard, and look at the *order* of the observations, which is mainly determined by the explanatory variables. Let $s_n = t_n^1 - t_n^0$ be the duration for the n^{th} individual and let $\iota(n)$ be the index of the person with the n^{th} shortest duration. In other words, for all $n < m$ we have

$$t_{\iota(n)}^1 - t_{\iota(n)}^0 = s_{\iota(n)} < s_{\iota(m)} = t_{\iota(m)}^1 - t_{\iota(m)}^0.$$

The partial likelihood is developed in the following way: consider the probability of individual i being the one with the shortest duration given that the shortest duration lasted t months and given the regressors x:

$$\mathscr{P}\left[\iota(1) = i | s_{\iota(1)}, x_1, x_2, \ldots, x_N\right] = \frac{\omega(x_i;\theta)}{\sum_{n=1}^N \omega(x_n;\theta)}.$$

This does not depend on the value of the baseline hazard function λ_0 at $s_{\iota(1)}$. The next step is to calculate the probability of individual j being the one with the second shortest duration given the length of the second shortest duration, the regressors x and given that individual i had the shortest duration:

$$\mathscr{P}\left[\iota(2) = j | s_{\iota(2)}, x_1, \ldots, x_N, \iota(1)\right] = \frac{\omega(x_j;\theta)}{\sum_{n \neq \iota(1)} \omega(x_n;\theta)}.$$

If we proceed in this fashion until all durations are exhausted and multiply the probabilities we can rewrite the result as

$$\mathscr{L}_p(\theta) = \sum_{n=1}^{N} \frac{\omega(x_n;\theta)}{\sum_{m \in R(s_n)} \omega(x_m;\theta)} \tag{5}$$

where the *risk set* R at t consists of those people who have durations no shorter than t:

$$R(t) = \{n = 1, 2, \ldots, N | s_n \geq t\}. \tag{6}$$

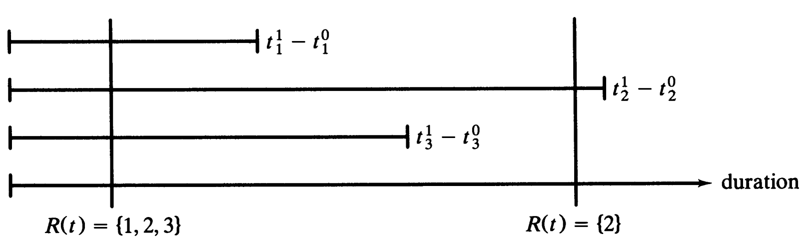
Figure 1 might give some intuition on the development of the likelihood function. The durations are ordered and then at each duration we consider the risk of exit for the person corresponding to that duration, relative to the people with durations exceeding that one.

Consistency and asymptotic normality of this estimator has been proven for various forms of this model. The most popular functional form for ω is log linear:

$$\omega(x;\theta) = \exp(\theta' x). \tag{7}$$

Tsiatis (1981) considered this case and gave a proof of the asymptotic properties based on convergence of the average score to a nonstochastic function. The complication in proofs of asymptotic normality lies in the fact that the scores are uncorrelated but not independent. Andersen and Gill (1982) extend the proof to the case where the characteristics x are allowed to vary over time. They give a proof in the context of counting processes. The main restriction on the covariate processes x is that they are *predictable* and *locally of bounded variation*. A sufficient condition

FIGURE 1.—RISK SETS FOR THE COX MODEL



for this to hold is that the covariate processes are continuous and have bounded first derivatives.

At this point another model deserves mention as it will later facilitate both the interpretation of the models with calendar-time dependence and the comparison with panel data analysis. In the fully parametric as well as the semi-parametric version of the models discussed so far, econometricians, Lancaster (1979), Heckman and Singer (1984), Meyer (1986), and Honoré (1990) among them, have been looking at the consequences of not observing some of the heterogeneity in the hazard function. The particular specification studied has been one where the unobserved heterogeneity enters multiplicatively. The hazard function for individual n, $\lambda^n(t)$ is assumed to be

$$\lambda^n(t) = \nu_n \cdot \lambda_0(t - t_n^0) \cdot \omega(x_n;\theta) \qquad (8)$$

with the unobserved ν_n independent of x_n and $\lambda_0(\cdot)$ parametrically specified or not. This modelling strategy is similar to the individual (random) effects approach in panel data analysis.

B. *Calendar-time Effects*

In this section the main ideas of the paper will be discussed. We will modify the Cox Regression Model introduced in the previous section in such a way that it allows for a flexible form of calendar-time dependence, at the expense of restricting the duration dependence to be parametrically specified. We will start by looking at alternative ways of incorporating dependence of the hazard on calendar time. The first possibility we look at is to add it as a time-varying regressor or *covariate* in the specification of the systematic part of the hazard rate:

$$\lambda(t, t - t^0, x) = \lambda_0(t - t^0) \cdot \omega(t, x;\theta). \qquad (9)$$

If we specify

$$\omega(t, x;\theta) = \exp \sum_{k=1}^{K} \theta_k h_k(t, x) \qquad (10)$$

with all functions h_k known, we are back in the log linear framework analyzed by Andersen and Gill (1982). Examples of such specifications are $h_1(t, x) = x$, $h_2(t, x) = \log(t)$ or $h_2(t, x) = t$. Another, more flexible, specification in the spirit of Ridder (1987) would be to define some $h_k(t, x) = I[c_k < t \leq c_{k+1}]$.[3] If $c_{k+1} - c_k$ were equal to one year, this would amount to introducing yearly dummies in the hazard function.

The second approach is to have as one of the regressors an indicator for the macroeconomic forces behind the nonstationarity of the model. In the case of employment and unemployment durations one might think of the national unemployment rate, the growth rate of GNP or a different indicator for the business cycle. In terms of the hazard function $\lambda^n(\cdot)$ for individual n:

$$\lambda^n(t) = \omega(u(t), t - t_n^0, x_n;\theta) \qquad (11)$$

or

$$\lambda^n(t) = \lambda_0(t - t_n^0) \cdot \omega(u(t), x_n;\theta) \qquad (12)$$

where $u(t)$ can be any variable that satisfies the regularity conditions for time-varying regressors, as given in Andersen and Gill (1982). In practice, these imply that $u(t)$ has to be a function of information available (just) before t. It does not, however, have to be strongly exogenous. Note that this type of covariate is different from the usual time-varying regressor. In this case the covariate is not individual-specific, and identification of its coefficients stems from the non-degenerate distribution of the entry date t^0 in the population. Flinn and Heckman (1983) follow this approach and use the national unemployment rate as the common, time-varying regressor. Lynch (1989) uses the local unemployment rate as a time-varying, but region-specific, regressor.

The above models, (4), (9), and (12), do not have all the disadvantages of the first, fully parametric model in (2). Nevertheless, they require the researcher to specify the dependence of the hazard on calendar time completely. The argument why this is difficult in practice is related to the reason behind including calendar time in the first place. The general justification for inclusion is that there might be forces behind the events that are both equal for everybody in the population as well as changing over time. An example is the life span of companies. Irrespective of their characteristics and the date the companies were set up, they might have correlated risks of going bankrupt at the same time via the phase of the business cycle, or seasonal fluctuations in demand. A similar story can be told for unemployment spells. If the general outlook is bad, the chances of finding a job might be slim for everybody, relative to the chances in good times. If this

[3] $I[.]$ is an indicator function, equal to one if the expression between the brackets is true, equal to zero otherwise.

is the case, one would ideally model these macro processes jointly with the individual behaviour or condition on their paths. However, this requires knowledge and observability of the exact processes that influence the hazard function. Because these processes might be characterized by high frequency movements (seasonals), low frequency movements (business cycles), as well as breaks (law changes), they might be difficult to approximate by low order polynomials. If not all these processes are observed, one should not restrict the time effect to a particular form. Techniques that do not require such a form are to be preferred.

Duration dependence, on the other hand, is an effect that can potentially be explained within economic models. Miller (1984), Jovanovic (1979), and Van Den Berg (1990) have studied models in which optimizing behaviour by individuals leads to hazard functions with particular forms of duration dependence. In the model studied by Van Den Berg, unemployed individual will lower their reservation wage over the unemployment spell in anticipation of a decline in benefit levels. This causes the hazard function to be an increasing function of duration. Jovanovic studies, among others, job-to-job transitions. He finds that the hazard should increase initially, when employee and employer learn about the quality (or the lack of quality) of the match, and then decrease, once the match has been found to be a successful one. These examples suggest that it is easier to model duration dependence than calendar time dependence. Specifically, they indicate that low order polynomials could capture most of the qualitative features of the duration dependence. In contrast, there is little reason to believe that low order polynomials could capture most of the high and low frequency changes in the time pattern of the hazard rate.

This is one of the reasons to propose reversing the roles of calendar time and duration in (9) as the third way to incorporate calendar time dependence. Instead of parametrizing the dependence on calendar time and leaving the dependence on duration free, one could specify the hazard in the following way:

$$\lambda(t, t-t^0, x) = \lambda_0(t) \cdot \omega(t-t^0, x; \theta) \quad (13)$$

with λ_0 an unknown function of time, and ω a known function. A second argument to prefer the semi-parametric specification (13) rather than any fully parametric specification is that we will still be able to utilize Cox's partial likelihood approach. Because this estimator requires only evaluation of the hazard rate, rather than evaluations of the integral of the hazard rate which are required by fully parametric procedures as exemplified in the likelihood function in (3), it can be computationally considerably easier. To underline the computational advantages of the estimator proposed in this paper, we will in the appendix outline how the estimates can be obtained using standard computer programs that estimate the conventional form of the Cox regression model with time-varying regressors.

Since (13) is the key equation we will discuss it in some detail before going on to the question of inference. $\lambda_0(\cdot)$ can represent trends in the hazard function, seasonals or business cycle effects. Its form can be of independent interest, depending on the particular application. Crucial is the assumption that its effect is proportional to that of the other variables.

The individual spells can now be correlated via the (potentially unobserved) common process $\lambda_0(t)$. They do not even have to be independent conditional on the realisation of this process, as long as (13) is the correct transition intensity if we condition on the realisations of the other histories up to t.

So far the similarity with the Cox Regression Model has been stressed. This similarity will also be used to derive an estimator for the proposed model. However, (13) can also be obtained via a different approach. Suppose we let the error term in the hazard specification (8) vary over time as well as vary over the population:[4]

$$\lambda^n(t) = \nu_{nt} \cdot \omega(t-t_n^0, x_n; \theta). \quad (14)$$

One needs more structure on this specification to be able to estimate parameters of interest. One way of providing such structure is to go back to (8) by restricting the error term to be constant over time: $\nu_{nt} = \nu_n$. Another possible way is to assume that the error term is not individual specific: $\nu_{nt} = \lambda_0(t)$. That would get us back to (13).

[4] Lancaster (1979) and later Heckman and Singer (1986) start from this general model, before making the assumption that the error term is constant over time (which in their case coincides with duration).

Another way of thinking about (13) might also provide some insights. The difference between (13) and (4) is one of time origins. In the original Cox model one uses the entry date as the time origin. In (13) we use a fixed calendar date as the time origin. The question of time origins has been posed before by Cox in a discussion of Oakes (1984). Kay (1982) applies a similar model to a multistage disease process. The time origin in his model is the start of the first stage. This implies that for the first stage duration and time coincide but for subsequent stages they are no longer the same. In his model duration is not used as an explanatory variable. A fixed calendar date and date of entry are two of the possible time origins. This does not exhaust the possibilities. If the hazard function also depends on age, one could use the birth date t^b as the time origin. An alternative model would in that case be:

$$\lambda(t, t^0, t^b, x) = \lambda_0(t - t^b) \cdot \omega(t, t - t^0, x; \theta). \tag{15}$$

For analyzing retirement decisions this might well be the appropriate time origin. This is also true in some medical applications, but there one would often interpret the birth date as the entry date, in which case age dependence would coincide with duration dependence. The following approach to inference for model (13) would go through in a similar way if one concentrated on age dependence.

Finally, it is worth remarking on the restriction embodied in (13). The time effect is assumed to be the same for everybody, multiplying the hazard function. One could relax this assumption by redefining some of the regressors and allowing them to be time dependent. For instance, if one wants to allow the effect of x to change over time, one could have two regressors, x and $x \cdot \ln t$, or x and $x \cdot t$. As long as the time dependence of the effects of regressors is parametrized, one can partial out the time dependence of the intercept, or the baseline hazard, $\lambda_0(t)$.

Going on to inference for the proportional hazards model (13), we start by assuming linearity of the logarithm of ω:

$$\omega(t - t^0, x; \theta) = \exp\left[\theta' h(t - t^0, x)\right] = \exp \sum_{k=1}^{K} \theta_k h_k(t - t^0, x). \tag{16}$$

Consider the ordered exit times $t^1_{\iota(1)} < t^1_{\iota(2)} < \cdots < t^1_{\iota(N-1)} < t^1_{\iota(N)}$, where $\iota(i)$ gives the identity of the i^{th} individual to exit. Conditional on all the regressors x, the entry dates t^0 and the first exist time $t^1_{\iota(1)}$, the probability of individual n being the first to exit is

$$\mathscr{P}\left[\iota(1) = n | t^1_{\iota(1)}, x_1, x_2, \ldots, x_N, t^0_1, t^0_2, \ldots, t^0_N\right]$$
$$= 0 \qquad \text{if } t^0_n \geq t^1_{\iota(1)}$$
$$= \frac{\exp\left[\theta' h\left(t^1_{\iota(1)} - t^0_n, x_n\right)\right]}{\sum_{m \in R(t^1_{\iota(1)})} \exp\left[\theta' h\left(t^1_{\iota(1)} - t^0_m, x_m\right)\right]} \qquad \text{if } t^0_n < t^1_{\iota(1)}$$

with the risk set $R(t)$ consisting of the people who entered but not exited before t.

$$R(t) = \left\{n = 1, 2, \ldots, N | t^0_n < t \leq t^1_n\right\}. \tag{17}$$

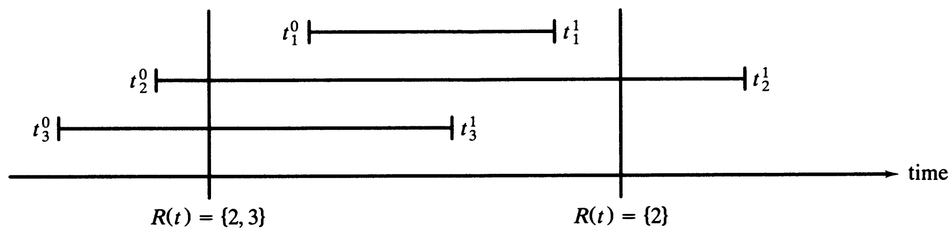
In the same manner we calculate the probability of individual m being the second to exit, given the identity of the first individual to exit, the regressors and entry times. If we proceed in this way until all the exit times are accounted for, we obtain the partial likelihood by multiplying all the probabilities:

$$\mathscr{L}_p(\theta) = \prod_{i=1}^{N} \mathscr{P}\left[\iota(i) | t^1_{\iota(i)}, x_1, x_2, \ldots, x_N, t^0_1, t^0_2, \ldots, t^0_N, (\iota(j), j < i)\right]$$
$$= \prod_{i=1}^{N} \frac{\exp\left[\theta' h\left(t^1_{\iota(i)} - t^0_{\iota(i)}, x_{\iota(i)}\right)\right]}{\sum_{m \in R(t^1_{\iota(i)})} \exp\left[\theta' h\left(t^1_{\iota(i)} - t^0_m, x_m\right)\right]}$$
$$= \prod_{n=1}^{N} \frac{\exp\left[\theta' h\left(t^1_n - t^0_n, x_n\right)\right]}{\sum_{m \in R(t^1_n)} \exp\left[\theta' h\left(t^1_n - t^0_m, x_m\right)\right]} \tag{18}$$

where the risk set is that defined in (17). In figure 2 the development of the likelihood function is given in such a way as to contrast it with that of the Cox likelihood function given in figure 1. Here the risk set varies over time whereas before it varied over duration.

To interpret the difference between the two versions of the Cox model, (4) and (13), which is at the core of this paper, in a different way, consider the following model that incorporates

FIGURE 2.—RISK SETS IN THE TIME-VARYING MODEL



both as special cases:

$$\lambda(t, t^0, x) = \lambda_0(t) \cdot \lambda_1(t - t^0) \cdot \omega(x; \theta). \tag{19}$$

By conditioning on the (ordered) durations $t^1 - t^0$, one at a time, we can get around specifying $\lambda_1(\cdot)$ and be flexible with respect to the duration dependence. If, on the other hand, we condition on the (ordered) exit times t^1, one at a time, we can avoid specification of $\lambda_0(\cdot)$ and be flexible with respect to calendar time dependence. Whichever we choose, if we want to exploit the computational advantages of the partial likelihood approach, we can be flexible with respect to only one of the two forms of time dependence,[5] at the expense of specifying the other up to a finite dimensional parameter vector. This is not to say that a completely non-parametric approach is not possible. From the data we can learn the distribution function $F(t|x, t^0)$ of t^1 given x and t^0 as long as there is enough variation in t^0 and x. From that we can calculate the hazard rate as a function of time, entry date and regressors as minus the derivative of the logarithm of one minus the distribution function:

$$\lambda(t, t^0, x) = -\frac{\partial}{\partial t} \ln\left[1 - F(t|t^0, x)\right].$$

Writing it as a function of duration and time rather than entry date and time amounts to an innocuous change of variable. The hazard function is therefore clearly identified non-parametrically, and it can be estimated by smoothing methods as kernel estimation. This leads to different rates of convergence and given that there are $\dim(x) + 2$ arguments in the hazard function such an approach would require a very large dataset.

Another approach would be to estimate a model such as (19) with both duration and time dependence free as a single index model. Such models typically require the potentially quite restrictive assumption that there is at least one regressor that varies continuously over the real line. In addition single index models would have to be adapted to allow for time-varying regressors and censoring that depends on regressors.

The above derivation of the partial likelihood based on specification (13) can easily be adapted to allow for censoring. Let the typical observation be (t^0, t^1, d, x), where $d = 1$ if t^1 is an exit or failure time, and $d = 0$ if the observation is censored at t^1 and exit is known to have not occurred at or before t^1. The appropriate partial likelihood function is obtained by multiplying the probabilities only for the proper exit times:

$$\begin{aligned} \mathscr{L}_p(\theta) &= \prod_{i=1}^{N} \mathscr{P}\left[(\iota(i)|t^1_{\iota(i)}, x, t^0, (\iota(j), j < i)\right]^{d_{\iota(i)}} \\ &= \prod_{n=1}^{N} \left[\frac{\exp[\theta' h(t_n^1 - t_n^0, x_n;)]}{\sum_{m \in R(t_n^1)} \exp[\theta' h(t_n^1 - t_m^0, x_m)]}\right]^{d_n} \end{aligned} \tag{20}$$

with the definition of the risk set $R(t)$ unchanged from (18). Its interpretation is now the set of people known to be at risk at t, i.e., people who have not exited or been censored before t. In the appendix sufficient conditions are given for the following result. They mainly involve regularity assumptions.

THEOREM 1: *Suppose that the hazard function for the individual with characteristics x whose entry date is t^0 is equal to $\lambda_0(t) \cdot \omega[\theta^{*\prime} h(t - t^0, x)]$. Suppose furthermore that assumptions* 1–4 *are satisfied. Let $\hat{\theta}$ be the maximand of $\mathscr{L}_p(\theta)$, with*

[5] If we treat age dependence as a third form of time dependence, this should read "we can be flexible with respect to only one of the three forms of time dependence."

$\mathscr{L}_p(\theta)$ given in (20). Then, as $N \to \infty$,

1. $\hat{\theta} \xrightarrow{p} \theta^*$
2. $\sqrt{N}(\hat{\theta} - \theta^*) \xrightarrow{d} \mathscr{N}(0, \Sigma(\theta^*)^{-1})$ *where* $\Sigma(\theta^*)$ *can consistently be estimated by*

$$\hat{\Sigma}(\theta) = \left[-\frac{1}{N} \frac{\partial^2 \mathscr{L}_p}{\partial\theta\, \partial\theta'}(\theta) \right]^{-1}$$

evaluated at $\theta = \hat{\theta}$.

Proof: see appendix.

Later we will be interested not only in the value of θ^* but also in that of the baseline hazard, $\lambda_0(t)$. It is difficult to estimate this directly. However, an estimator for the integrated baseline hazard can be constructed in the same way as proposed by Breslow (1972, 1974) for the original Cox model. Define:

$$\Lambda(t, \underline{t}) = \int_{\underline{t}}^{t} \lambda_0(s)\, ds.$$

This integrated hazard has an interpretation that is different from its counterpart in the Cox model. There it is the integrated hazard for the typical individual with $x = 0$. Here there is no individual for whom it is the integrated hazard since it leaves out the duration dependence which cannot be kept constant. It does have an interpretation though, as an indicator for the hazard an individual faces if he enters at that particular date.

The Breslow estimator for $\Lambda(t, \underline{t})$ is

$$\hat{\Lambda}(t_j^1, \underline{t}) = \sum_{\underline{t} < t_n^1 \le t_j^1} \frac{1}{\sum_{m \in R(t_n^1)} \exp\left[\hat{\theta}' h(t_n^1 - t_m^0, x_m)\right]} \tag{21}$$

for $j = 1, 2, \ldots, N$, and linear interpolation for values of t in between two exit times.

The partial likelihood approach outlined above formally requires continuous time data. In practice, data are always discrete. This causes complications only if the grid is so coarse that there are a large number of ties. To deal with ties we use one of the approximations proposed by Cox and Oakes (1984), which is consistent with the last representation of the partial likelihood function in (18). If $t_n^1 = t_m^1$ for two observations $m \neq n$, the likelihood contribution of observations n and m is

$$\frac{\exp\left[\theta' h(t_n^1 - t_n^0, x_n)\right]}{\sum_{l \in R(t_n^1)} \exp\left[\theta' h(t_n^1 - t_l^0, x_l)\right]} \times \frac{\exp\left[\theta' h(t_m^1 - t_m^0, x_m)\right]}{\sum_{l \in R(t_m^1)} \exp\left[\theta' h(t_m^1 - t_l^0, x_l)\right]}$$

$$= \frac{\exp\left[\theta' h(t_n^1 - t_n^0, x_n) + \theta' h(t_m^1 - t_m^0, x_m)\right]}{\sum_{l \in R(t_n^1)} \exp\left[\theta' h(t_n^1 - t_l^0, x_l)\right]}$$

because $t_m^1 = t_n^1$ implies that $R(t_n^1) = R(t_m^1)$. If the number of spells is very large relative to the number of periods, one might want to deal with the problem of ties in a more direct way. In that case one could specify the hazard in discrete time as

$$\lambda(t, t^0, x) = \lambda_{0l} \cdot \exp\left[\theta' h(t - t^0, x)\right] \quad \text{for } t \in [c_{l-1}, c_l), \quad \text{for } l = 1, 2, \ldots, L$$

with c_l known for $l = 0, 1, 2, \ldots, L$, and treat the λ_{0l} as a set of $L + 1$ unknown parameters which have to be estimated jointly with the K dimensional parameter θ. Such a procedure is computationally more demanding than the partial likelihood approach because it requires maximization over a $L + 1 + K$ dimensional space (84 + 3 in the application in the next section if we use monthly intervals) where the partial likelihood estimator only requires maximization over a K dimensional (3 in the application in the next section) space. The discrete time estimator has the advantage of immediately providing estimates of the baseline hazard. The discrete time approach is similar to the probit models with structural group effects studied by Borjas and Sueyoshi (1991). The appropriateness of the discrete time versus the continuous time approach depends on the relative number of spells versus the number of periods. In particular if there are no ties, and the number of spells is smaller than the number of periods, only the partial likelihood estimator is feasible.

III. An Application

In this section the methodological analysis of section II will be applied to a Dutch dataset on labor market histories. In section IIIA the dataset will be described and some summary statistics given. In the second subsection we will first analyze the employment to unemployment transition using the various models of section II and subsequently we will look at the other transitions.

A. The Data

The dataset used in this paper was a part of the ORIN (Onderzoek naar Relatievormen in Nederland) dataset. It was setup by NIDI (Nederlands Interuniversitair Demografisch Instituut) in cooperation with the Universities of Tilburg, Amsterdam and Wageningen.[6] The part of the dataset with which we are concerned contains labor market histories for at least seven years. A random sample of Dutch men (in this analysis we will only look at the labor market histories of men) were asked in 1984 to reconstruct their labor market histories going back to the last change in labor market status before January 1977. Three labor market states are identified: working full time, working part-time and not working. No distinction was made in the dataset between being unemployed while actively seeking work and being out of the labor force. Flinn and Heckman (1983) have shown that this distinction is potentially important. To avoid complications arising from the inability to distinguish between being unemployed and being engaged in full-time education or that between being unemployed and being retired, only the observations on people between 23 and 50 years of age in January 1977 were used. This leaves us with 372 labor market histories. For males in this age group part-time employment is a relatively rare phenomenon and for the purpose of this study the two categories of employment have been aggregated into one.

We will treat the dataset as if it were constructed by taking a random sample in January 1977 from the population of males between 23 and 50 years of age at that time, and following them till January 1984. This would be valid if the population of males between 23 and 50 years of age in January 1977 would coincide with the population of males between 30 and 57 years of age in January 1984. This is of course not exactly true. Mortality and emigration lead to an outflow on the one hand and immigration leads to an inflow on the other hand. In the application in this paper we ignore these effects, which are arguably small for males in this age group in the years 1977–1983 in the Netherlands. We condition on the labor market histories up to January 1977. In principle, there is information in the labor market histories before January 1977, but because the length of the observed labor market histories before January 1977 varies across the population, we would need stationarity assumptions to extract this for all the estimators, and even stronger assumptions for the partial likelihood estimator. We therefore do not use this information and condition on events before January 1977. The dataset contains monthly data, whereas we have been assuming continuous data in the preceding part of the paper. However, due to the relatively small number of ties we did not think this would lead to a serious bias in the results. Using monthly dummies is in principle possible, but since there are many months without any transition, there would be serious doubts about the accuracy of such estimates.

Given the two states, employed (E) and not employed (U) (or unemployed as both terms will be used interchangeably in the remainder of this paper), three transitions are distinguished. Someone who is unemployed can move into employment. Someone who is employed can either change jobs, which will be considered an employment to employment transition or he can move into unemployment. The three transition types will be denoted by UE, EE and EU, respectively.

Two explanatory variables are used. The first one is an individual's age in 1977. For ease of computation age is not used as a time-varying regressor. Theoretically there is no problem with treating it either way. The second explanatory variable is an index for education. It ranges from 1 to 5. The higher values indicate higher levels of education. In table 1 summary statistics are given for the regressors in the sample of 372 men.

The power to estimate transition rates obviously comes from the occurrence of transitions in the sample. Table 2 gives the marginal distribu-

TABLE 1.—SUMMARY STATISTICS EXPLANATORY VARIABLES

Variable	Mean	S.D.
Age	34.9	7.2
Education	2.8	1.1

[6] I wish to thank the NIDI for making these data available to me.

TABLE 2.—FREQUENCIES OF TRANSITIONS

Type	*UE*	*EU*	*EE*	All
0	321	246	164	101
1	46	117	159	147
2	3	8	39	80
3	2	1	9	27
≥ 4	0	0	1	17

tion of the number of transitions of the different types.

It shows, for example, that there are 321 men with no transitions from unemployment into employment. There are 46 men with one such transition, 3 with two transitions from unemployment into employment, 2 with three such transitions, and nobody with four or more transitions from unemployment into employment.

The total number of unemployment to employment transitions is 51. The total number of employment to unemployment transitions is 126, leaving a net flow into unemployment of 75. Partly this is a reflection of the aggregate conditions in the Dutch economy in those years, and this is of course the effect we are trying to incorporate in our analysis. Partly it might also be due to the fact that as a group, the sample grows seven years older, leading to a higher expected proportion of non-employed for various reasons (early retirement,[7] permanent sick leave).

First we look at estimates of the hazard by year to investigate whether there are obvious trends and patterns over time in the hazard rates. Suppose the hazard function for the transition from state i to state j, given that state i was entered at t_0 has the form:

$$\lambda^{ij}(t, t^0, x) = \lambda^{ij}_k \quad \text{for } c_{k-1} < t \leq c_k, \qquad k = 1, 2, \ldots, K$$

where $c_0 = 0$ and $c_K = \infty$, for the transition from state i to state j. This model is very close to the piecewise constant hazard discussed by Cox and Oakes (1984) and Lancaster (1990). The difference is in the time scale that is calendar time here, and duration in their models. Ridder (1987) gives comparable summary statistics. Table 3 gives the estimates of λ^{ij}_k for yearly intervals ($c_k - c_{k-1} = 1$ year) for the three transitions.

The transition rates are not very precisely estimated. Nevertheless, there is some information in the table. If we look at the last column, where the national unemployment rate[8] is given, one can see that the four highest unemployment figures occurred in the last four years. This coincides with the four highest *EU* transition rates and the four lowest *EE* transition rates. More formally, if we look at the Spearman rank correlation coefficients, we find 0.70, 0.21, and −0.68 for the correlation between the national unemployment rate and the *EU*, *UE* and *EE* transition rates, respectively.[9]

B. *The Results*

In this subsection we will apply the techniques developed in section II to the dataset described in section IVA. Initially we will concentrate on the employment to unemployment transition to show the full force of the techniques. In the second half of this section we will turn to the other two transitions. Before this we will give the full likelihood function on which the estimations are based, either directly, or indirectly via the associated partial likelihood function. Let $Y^n_E(t) = 1 - Y^n_U(t)$ be equal to 1 if individual n is in state E (is employed) at t and 0 when he is in state U (not employed) at t, for $t \in [0, 84]$, where t is in months from January 1977 onwards. Let

$$A^n_i = \{t \in [0, 84] | Y^n_i(t) = 1\}$$

be the union of all spells in state i, for $i = E, U$ by individual n. Let $N^n_{ij}(t)$ be the number of i to j transitions between January 1977 and t for this individual, for $t \in [0, 84]$, for $ij \in \{EU, UE, EE\}$. $s_n(t)$ is equal to t minus the date of the last

[7] We expect to have reduced this to a minimum by restricting the sample to under fifty year olds.

[8] Registered male unemployment, as a percentage of the dependent labor force, from the *SociaalEconomische Maandstatistiek*, a Central Bureau of Statistics publication.

[9] The definition of the unemployment rate in table 3 does not correspond exactly to the definition of not employed in our sample. The main reason is that being not employed does not imply being registered unemployed. The latter is necessary in order to qualify for benefits but not otherwise compulsory. Even though the levels of sample non-employment rate and population unemployment rate do not match closely, the correlation is high (0.92 for levels and 0.89 for logarithm), and we will use the national registered unemployment rate as a proxy for the national equivalent of the proportion of not employed in the sample.

TABLE 3.—YEARLY TRANSITION AND UNEMPLOYMENT RATES

Year	*EU*		*UE*		*EE*		*U*-rate
1977	.030	(.010)	.052	(.040)	.052	(.012)	4.6
1978	.022	(.008)	.121	(.054))	.049	(.012)	4.3
1979	.028	(.010)	.041	(.029)	.043	(.012)	4.1
1980	.071	(.016)	.065	(.032)	.025	(.010)	5.1
1981	.083	(.017)	.037	(.022)	.031	(.011)	8.2
1982	.080	(.017)	.094	(.031)	.036	(.012)	11.6
1983	.053	(.014)	.102	(.031)	.026	(.010)	16.6

transition before t, or the time spent in the current spell. At $t = 0$ this equals time spent in the initial state, since we do have information on that.

The likelihood function is now:

$$\mathscr{L}(\theta) = \mathscr{L}^{EE}(\theta_{EE}) \cdot \mathscr{L}^{EU}(\theta_{EU}) \cdot \mathscr{L}^{UE}(\theta_{UE})$$

with, for $ij \in \{EE, EU, UE\}$,

$$\mathscr{L}^{ij}(\theta_{ij}) = \prod_{n=1}^{N} \prod_{k=1}^{N_{ij}^n(84)} \lambda_{0ij}(t_{ijk}^n) \times \omega\left(s_n(t_{ijk}^n), x_n; \theta_{ij}\right) \times \exp\left[-\int_{A_i^n} \lambda_{0ij}(u) \times \omega\left(s_n(u), x_n; \theta_{ij}\right) du\right].$$

We estimate three parametrizations of the transition model. The first one does not allow for calendar time dependence. The second one allows for a parametrized form of time dependence, where the parametrization uses a common time varying regressor. The third model uses the modified Cox regression model where the time dependence is not parametrically specified. The three models are nested which allows one to see clearly how the results change with the level of sophistication in the time dependence specification. The functional form of the ω part of the hazard function is the same for all three specifications:

$$\omega(s, x; \theta_{ij}) = s^{\theta_{ij1}} \cdot \exp(x_1\theta_{ij2} + x_2\theta_{ij3}).$$

This functional form allows the duration dependence to be negative or positive. Like the effect of the time invariant regressors, it is restricted to be monotone. The three models differ in the way the time dependence is treated:

Model 1. *No time dependence*:

$$\lambda_{0ij}(t) = \exp(\theta_{ij0}).$$

This model assumes that the durations have a Weibull distribution. This is a fairly flexible distribution that has widely been used in econometrics since Lancaster (1979). It will serve here as a benchmark case against which we can judge the more complicated models.

Model 2. *Parametric time dependence*:

$$\lambda_{0ij}(t) = \exp\left(\theta_{ij0} + \theta_{ij4} \cdot g(t)\right)$$

where $g(t) = \ln(u(t)/(1 - u(t)))$ with $u(t)$ the national unemployment rate at t. The second model attempts to capture the variation over time by putting in a time varying regressor. The choice of the particular regressor can be motivated as follows: suppose the transition rates from unemployment to employment and vice versa can be written as

$$\lambda^{EU}(t) = \exp[\beta_0 + \beta_1 \cdot h(t)]$$

and

$$\lambda^{UE}(t) = \exp[\alpha_0 + \alpha_1 \cdot h(t)]$$

for some possibly unknown or unobserved macro process $h(t)$. The equilibrium unemployment rate u^* associated with h is that level of unemployment for which the flows into and out of employment are equal:

$$\lambda^{EU}(h) \cdot (1 - u^*(h)) = \lambda^{UE}(h) \cdot u^*(h).$$

Over time there is a one-to-one correspondence between $h(t)$ and $u^*(t)$:

$$h(t) = \frac{1}{\beta_1 - \alpha_1} \cdot \left[\alpha_0 - \beta_0 + \ln \frac{u^*(t)}{1 - u^*(t)}\right].$$

Hence, λ^{EU} and λ^{UE} are also loglinear in

$\ln[u^*/(1 - u^*)]$. If the actual unemployment rate is a good approximation for the equilibrium one, using the first as a regressor might be a good alternative to using $h(t)$ if the latter is hard to pin down or measure. The coefficient on this regressor should not be given too much interpretation. We are not trying to explain the individual transition probabilities with the average or national one. What we are trying to achieve by putting in this regressor is to control for changes in the average transition rate when estimating the effect of individual specific explanatory variables. Because we do not have the national proportion of not employed we use the national rate of registered unemployment. Flinn and Heckman (1983) estimate almost the same model. Their specification for $\omega(\cdot)$ is the same as ours, and the logarithm of λ_{0ij} is assumed to be linear in the national unemployment rate (rather than in the logarithm of $u(t)/[1 - u(t)]$ as in our specification).

Model 3. *Semiparametric time dependence*: $\lambda_{0ij}(t)$ is not specified. The third model finally estimates the effect of duration and explanatory variables while leaving the time dependence free. In nests the preceding two models. The coefficient on any macro process (like the unemployment rate in the second model) is no longer identified. This is an important aspect of the distinction between individual time varying regressors and common time varying regressors, alluded to earlier in section IIB.

There is a difference in the interpretation of the coefficient for duration, θ_{ij1} between model 1 and models 2 and 3. In model 1, the conventional interpretation of this coefficient is that it indicates how the transition rate would change if someone remained in his current state for a longer period of time. Formally, it is the logarithm of the ratio of the hazard he would face $(e - 1) \times (t - t^0)$ time units from now and the hazard he currently faces. This comparison of risks faced by the same person at different times does not work in models 2 and 3 because the hazard rate changes with time and duration. In that case θ_{ij1} has to be interpreted as an indicator of the difference between the risks faced at the same time by people who have the same time invariant characteristics, one of whom has been in his current state for a longer period than the other. This interpretation does work in all three models and is therefore the preferred one for comparisons of the three models.

Table 4 gives the estimation results for models 1–3 for the employment to unemployment transition.

A couple of remarks on the results will be made. The log likelihood of the third model is not comparable to that of the other two models because it is based on a partial likelihood. In general one can evaluate models like this where one has a series of nested models in two ways. One can either look at that part of the model with is common to all, or one can look at that part of the model which is different for every model. The first, common, part of the models is that relating the hazard to duration and time invariant regressors. The coefficients on the static regressors do not change very much with the level of sophistication with which we treat the time dependence. The coefficient on duration, however, does change considerably. In the most general model, the effect is significantly different from zero at the 90% level. Going from model 1 to 2 it is clear that there is time dependence. The coefficient on the unemployment rate is significantly different from zero, and the log likelihood goes up considerably. A test for the time invariance of the parameters in Model 1 was performed. An extra regressor t was added. The score function for this case is

$$\mathscr{S}(t^1, t^0, x; \theta) = t^1 - \int_{t^0}^{t^1} s \cdot \omega(s - t^0, x; \theta)\, ds.$$

TABLE 4.—THE *EU* TRANSITION

Variable	Model 1		Model 2		Model 3	
Intercept	−5.07	(0.46)	−5.38	(0.49)	—	—
Log Duration	−0.07	(0.11)	−0.09	(0.11)	−0.17	(0.10)
Age	0.03	(0.01)	0.03	(0.01)	0.02	(0.01)
Education	−0.17	(0.10)	−0.18	(0.10)	−0.19	(0.10)
Unempl. rate	—	—	0.47	(0.18)	—	—
Log Likelihood	−688.0		−685.1		−604.3	

The value of the test statistic was 16.2. As the 95% quantile of the appropriate $\chi^2(1)$ distribution is 3.6, it is clear that we can reject the null hypothesis of no time dependence.

One way to evaluate Model 2 versus Model 3 is to perform a Hausman test on the common parameters. This is a test of the hypothesis that $\lambda_0(t) = \exp(\theta_4 \cdot \ln(u(t)/(1-u(t))))$ for some parameter θ_4. Since the two models are nested, and since the estimates for Model 2 are efficient if the null hypothesis is correct, the difference in the parameter estimates asymptotically has a normal distribution with mean zero and variance equal to the difference in the variances. Under the null hypothesis the variance of the coefficients should be larger under Model 3 than under Model 2. One can see immediately from table 4 that this is not true for the coefficient on log duration. This is evidence against the null hypothesis.

The second way of evaluating the relative performance of the three models is by looking at part where they differ, the baseline hazard, $\lambda_0(t)$, or equivalently, the integrated baseline hazard, $\Lambda(t,0)$. For the three models this is estimated as

Model 1. *No time dependence*:

$$\hat{\Lambda}(t,0) = t \cdot \exp(\hat{\theta}_0).$$

Model 2. *Parametric time dependence*:

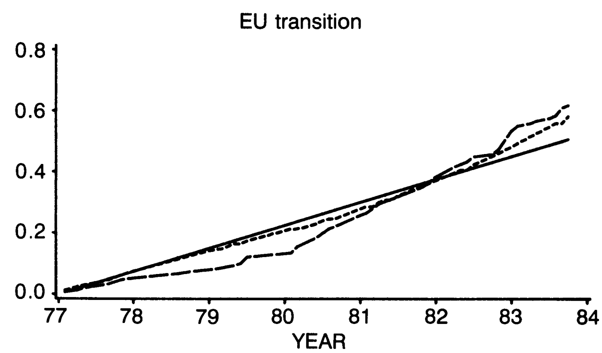
$$\hat{\Lambda}(t,0) = \int_0^t \exp\Big(\hat{\theta}_0 + \hat{\theta}_4 \times \ln[u(s)/(1-u(s))]\, ds\Big).$$

Model 3. *Semi-parametric time dependence*:

$$\hat{\Lambda}(t,0) = \sum_{n=1}^{N} \sum_{k=1}^{N^n(t)} \left[\sum_{m \in R_E(t_k^n)} \exp\Big[\hat{\theta}_1 s_m(t_k^n) + \hat{\theta}_2 x_{1m} + \hat{\theta}_3 x_{2m}\Big]\right]^{-1}.$$

The solid, dotted and dashed lines in figure 3 correspond to models 1, 2, and 3, respectively. It clearly shows that the slope of the integrated hazard is not constant over time. It increases markedly during 1980. This figure makes the interpretation of the bias in the duration dependence easier. Over time the average hazard increases. If one, incorrectly, does not take this into account, it appears that the hazard does not decrease very rapidly with duration. The increase with calendar time and the decrease with duration partially cancel each other out if they are not both incorporated in the model.

FIGURE 3.—INTEGRATED BASELINE HAZARD FOR THE *EU* TRANSITION



The figure also suggests an alternative test of the model. One could test whether all coefficients change at this point in time where the intercept seems to change so dramatically. To implement this test we first split the labor market histories into two parts, one describing the first three and a half years and the second giving the subsequent three and a half years. We then estimate model 1 for the two parts separately without any restrictions. Note that if the only time varying coefficient were the intercept, and if this only changes in July 1980, then both this split sample estimation, and the estimates of model 3 would be consistent. In table 5 the Weibull model is estimated for the two subperiods. These coefficients are not very precisely estimated. A formal test of the hypothesis that all coefficients are the same across both periods gives a test statistic of 26.6, where the 95% quantile of a $\chi^2(4)$ distribution is 9.5. If we test the equality of the last three coefficients, i.e., all coefficients but the constant term, we get a test statistic equal to 1.7. The 95% quantile of the appropriate $\chi^2(3)$ distribution is 7.4. We can therefore not reject the hypothesis that only the intercept varies with time.

Now we will turn to the other two transitions. The same three models are estimated with $\omega(\cdot)$ log linear in age, education and the logarithm of duration, and the time dependence going from non-existing via parametrically specified to semi-parametric. For both transitions the integrated baseline hazard was calculated. In table 6 and figure 4 the results for the unemployment to employment transition are given.

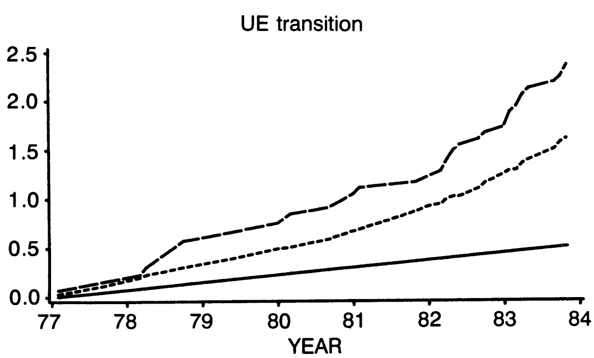
TABLE 5.—WEIBULL MODEL FOR TWO SUBPERIODS FOR THE *EU* TRANSITION

	Jan 1977 to June 1980		July 1980 to Dec 1983	
Variable	Coefficient	S.D.	Coefficient	S.D.
Intercept	−6.73	(1.07)	−4.29	(0.59)
ln Duration	0.11	(0.26)	−0.18	(0.14)
Age	0.06	(0.03)	0.02	(0.02)
Education	−0.19	(0.23)	−0.16	(0.11)
Log Likelihood	−166.6		−411.8	

TABLE 6.—THE *UE* TRANSITION

	Model 1		Model 2		Model 3	
Variable	Coefficient	S.D.	Coefficient	S.D.	Coefficient	S.D.
Intercept	−3.94	(0.53)	−2.23	(0.82)	—	—
Log duration	−0.24	(0.17)	−0.27	(0.17)	−0.38	(0.17)
Age	−0.08	(0.02)	−0.09	(0.02)	−0.09	(0.03)
Education	0.31	(0.15)	0.32	(0.15)	0.31	(0.16)
Unempl. rate	—	—	0.68	(0.31)	—	—
Log likelihood	−206.4		−203.9		−141.9	

FIGURE 4.—THE INTEGRATED BASELINE HAZARD FOR THE *UE* TRANSITION



The estimates of the coefficients on the time invariant regressors are again not much affected by the level of sophistication with which the time dependence is treated. The duration dependence is remarkably negative for the last model. After 6 months the value of the time invariant part of the hazard, ω, has decreased to 50% of its level after one month. Again this strong duration dependence is partly obscured in the models with more tightly specified dependence or no time dependence at all. Again a Hausman test leads to rejection of Model 2 in favor of Model 3 because the difference in variances is not positive definite.

In table 7 and figure 5 the results for the job to job transition are shown.

For this transition neither the estimates for duration dependence nor those for the effects of the time invariant regressors are seriously affected by ignoring the time dependence. From figure 5 it can be seen that the increased flexibility does not change the estimates considerably. This time the Hausman test for equality of the coefficients under Models 2 and 3 gives a value of 2.5. Given that the asymptotic distribution of the tests statistic is chi-squared with three degrees of freedom, we cannot reject the hypothesis that the variation over time is captured by the variation in the national unemployment rate.

Now we will turn to another aspect of the models with time dependence. The piecewise constant hazard estimates in table 3 can be interpreted as the average hazard over the yearly intervals. If we look at this average at a particular point in time, we can decompose it into two factors:

$$\bar{\lambda}_{ij}(t) = E_{s,x}\big[\lambda_{0ij}(t) \times \omega\big(s(t), x; \theta_{ij}\big) | Y_i(t) = 1\big] = \lambda_{0ij}(t) \cdot \bar{\omega}_{ij}(t)$$

with

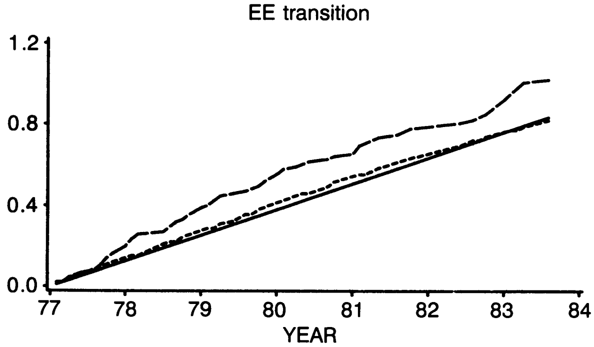
$$\bar{\omega}_{ij}(t) = \int \omega(s, x; \theta_{ij}) \cdot f_i(s, x | t)\, ds\, dx$$

where $f_i(s, x|t)$ is the distribution of (incomplete) durations s and characteristics x at time t among

TABLE 7.—THE *EE* TRANSITION

Variable	Model 1 Coefficient	Model 1 S.D.	Model 2 Coefficient	Model 2 S.D.	Model 3 Coefficient	Model 3 S.D.
Intercept	−4.55	(0.52)	−5.22	(0.85)	—	—
Log duration	−0.26	(0.13)	−0.25	(0.13)	−0.29	(0.14)
Age	0.03	(0.02)	0.03	(0.02)	0.03	(0.02)
Education	−0.05	(0.10)	−0.05	(0.10)	−0.04	(0.10)
Unempl. rate	—	—	−0.24	(0.22)	—	—
Log likelihood	−532.1		−531.4		−448.5	

FIGURE 5.—THE INTEGRATED BASELINE HAZARD FOR THE *EE* TRANSITION



that part of the population that is in state i at that time. $\bar{\lambda}_{ij}$ changes over time because the baseline hazard λ_{0ij} changes or because the distribution f_i changes. The changes in the baseline hazard have been shown in Figures 3, 4, and 5. It is interesting to contrast that with the effects of the changes in the population distribution of the explanatory variables. We can estimate the $\bar{\omega}_{ij}(t)$ with

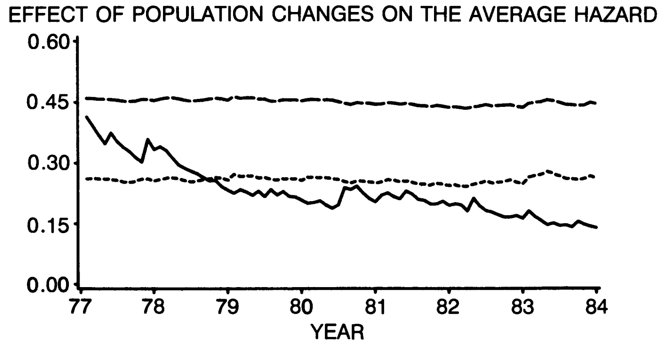
$$\hat{\omega}_{EE}(t) = \sum_{n=1}^{N} Y_E^n(t) \times \omega(s_n(t), x_n; \hat{\theta}_{EE}) \Big/ \sum_{n=1}^{N} Y_E^n(t)$$

$$\hat{\omega}_{EU}(t) = \sum_{n=1}^{N} Y_E^n(t) \times \omega(s_n(t), x_n; \hat{\theta}_{EU}) \Big/ \sum_{n=1}^{N} Y_E^n(t)$$

$$\hat{\omega}_{UE}(t) = \sum_{n=1}^{N} Y_U^n(t) \times \omega(s_n(t), x_n; \hat{\theta}_{UE}) \Big/ \sum_{n=1}^{N} Y_U^n(t).$$

In figure 6 the paths of $\hat{\omega}_{EE}(t)$ (the dotted line), $\hat{\omega}_{EU}(t)$ (the dashed line), and $\hat{\omega}_{UE}(t)$ (the solid line) are given. It can be seen that a large part of the change in the average *UE* hazard is due to the change in the distribution of x and s among the unemployed. In the seven years the characteristics of the unemployed change consid-

FIGURE 6.—CHANGES IN THE CONDITIONAL HAZARD OVER TIME



erably. Increasingly they are people with low chances of finding a job again even in good economic times when the baseline hazard is high. They are characterized by having been unemployed for a considerable period already, and/or having low levels of education, and/or being relatively old. On the other hand, the distribution of characteristics among those employed does not change much. There the change in the average hazard is mainly due to the change in the baseline hazard.

The overall story for the increase in unemployment over the years 1977–1983 emerging from this analysis can now be summarized as follows: the unemployment to employment baseline hazard was relatively low between 1979 and 1982, leading to lower chances for everybody to escape from unemployment. After that year the baseline hazard increased again, but the positive effect of that was countered by two negative effects. The employment to unemployment baseline hazard had increased sharply in 1980, leading to a higher flow into unemployment. Also, the people who were unemployed at that time had longer histories of being unemployed, leading to lower chances of escaping from this situation. These two effects more than outweighed the effect of the increase in the baseline hazard from unemployment to employment after 1982, ensuring that the unemployment rate continued its upward path.

IV. Conclusion

In this paper an alternative form of the proportional hazard model is proposed. Instead of leaving the duration dependence unspecified, we leave the calendar time dependence unspecified and parametrize the duration dependence. This allows one to take account of the effects of a changing macroeconomic environment on the durations of jobs. Since most duration data contain information about periods covering a number of years, the assumption of a stationary environment, while often made, is unappealing. This model can be viewed as a modification or alternative interpretation of the Cox regression model. In the conventional form of the Cox model, we measure time from the date of entry. In this paper we use a fixed calendar date as the time origin. Inference proceeds along similar lines as that for the original Cox model. When we apply this model to Dutch data on labor market histories the results change markedly compared to those with conventional models. Ignoring calendar time dependence severely biases the estimates of duration dependence. Incorporating these calendar time effects by means of a common, time-varying regressor as the unemployment rate is not sufficient to eliminate this bias. Estimates of the effects of time invariant regressors do not seem to be seriously affected. It is shown that the variation over time has been a major factor in the increase of the unemployment rate in the Netherlands in the late seventies and early eighties.

APPENDIX A

Proof of Theorem 1

In this appendix we will give sufficient conditions for theorem 1. First we will change notation slightly. This will enable us to use results by Andersen and Gill (1982).

Define $N^n(t)$, $Y^n(t)$ and $s_n(t)$:

$$N^n(t) = I\left[t_n^1 \leq t\right] \cdot I[d_n = 1]$$

$$Y^n(t) = I\left[t_n^0 < t\right] \cdot I\left[t \leq t_n^1\right]$$

$$s_n(t) = t - t_n^0 \text{ if } t \geq t_n^0 \text{ and 0 otherwise.}$$

$N(t)$ counts the number of events of interest between 0 and t. $Y(t)$ indicates whether the individual is at risk at time t.

We assume

$$\lim_{u \uparrow t} \lim_{dt \downarrow 0} \mathscr{P}\left[N^n(u+dt) - N^n(u) = 1 \middle| \{N(v), Y(v)\}_{v \leq u}, x\right] / dt = Y^n(t) \cdot \lambda_0(t) \cdot \exp\left[\theta^{*\prime} h(s_n(t), x)\right]$$

with $Y(\cdot)$ and $s(\cdot)$ predictable processes.

The partial likelihood function is:

$$\mathscr{L}_p(\theta) = \prod_{n=1}^{N} \prod_{k=1}^{N^n} \frac{Y^n(t_k^n) \cdot \exp\left[\theta' h\left(s_n(t_k^n), x_n; \theta\right)\right]}{\sum_{m=1}^{N} Y^m(t_k^n) \cdot \exp\left[\theta' h\left(s_m(t_k^n, x_m; \theta\right)\right]}. \tag{22}$$

Define $\hat{\theta}_N$ to be the maximal of (40).

Assumption 1: $Y^n(t) = 0$ *if* $t < 0$ *or* $t > b$. $x \in X$, *a compact subset of* $\mathscr{R}^L$. $\theta \in \Theta$, *a compact subset of* $\mathscr{R}^K$

Assumption 2: $0 \leq \lambda_0(t) \leq c$ *for all* $t \in [0, b]$. $h(s, x)$ *is continuous on* $(0, b] \times X$.

To ensure identification and asymptotic normality one has to look at the first and second moments of h. Define:

$$S^{(0)}(\theta,t)=\frac{1}{N}\sum_{n=1}^{N}Y_n(t)\cdot\exp[\theta' h(s_n(t),x_n)]$$

$$S^{(1)}(\theta,t)=\frac{1}{N}\sum_{n=1}^{N}h(s_n(t),x_n)\cdot Y_n(t)\cdot\exp[\theta' h(s_n(t),x_n)]$$

$$S^{(2)}(\theta,t)=\frac{1}{N}\sum_{n=1}^{N}h(s_n(t),x_n)\cdot h(s_n(t),x_n)'\cdot Y_n(t)\exp[\theta' h(s_n(t),x_n)]$$

and

$$\Sigma(\theta)=\int_0^b\left[\frac{S^{(2)}(\theta,t)}{S^{(0)}(\theta,t)}-\frac{S^{(1)}(\theta,t)}{S^{(0)}(\theta,t)}\cdot\frac{S^{(1)}(\theta,t)'}{S^{(0)}(\theta,t)}\right]\cdot S^{(0)}(\theta,t)\lambda_0(t)\,dt.$$

Assumption 3: $S^{(0)}$, $S^{(1)}$ and $S^{(2)}$ converge to their expectation uniformly in θ and t. $\Sigma(\theta^*)$ is positive definite.

Assumption 4: For all $\epsilon>0$:

$$\sup_x\int_0^b\|h(s,x)\|\cdot I[N^{-1/2}\cdot\|h(s,x)\|>\epsilon]\exp[\theta' h(s,x)]\,ds\xrightarrow{p}0.$$

The last assumption replaces condition C in Andersen and Gill (1982). In both cases it is trivially fulfilled if h is bounded. The advantage of the formulation here is that the Weibull specification is included. To see this, let $h(s,x)=\ln(s)$, and assume that the coefficient is $\theta>-1$. Then,

$$\int_0^b\|\ln(s)\|\cdot I[N^{-1/2}\|\ln(s)\|>\epsilon]\exp[\theta\ln(s)]\,ds=-\int_0^{a(N)}\ln(s)\cdot s^{\theta}\,ds$$

for $a(N)=\exp[-N^{1/2}\cdot\epsilon]$. This is equal to

$$-\int_0^{a(N)}\ln(s)\cdot s^{1/2+\theta/2}\cdot s^{\theta/2-1/2}\,ds$$

which is, for N large enough, and therefore $a(N)$ small enough, bounded by

$$\int_0^{a(N)}s^{\theta/2-1/2}\,ds.$$

This goes to zero as $a(N)$ goes to zero, which shows that assumption (5) is satisfied.

Proof of Theorem 1: We check the conditions for lemma 3.1 and theorem 3.2 in Andersen and Gill with the interval [0, 1] replaced by [0, b]. Conditions A, B and D follow trivially from our assumptions. Condition C is only necessary to guarantee that as $N\to\infty$,

$$\int_0^b\frac{1}{N}\sum_{n=1}^{N}\|h\|\cdot I[N^{-1/2}\cdot\|h\|>\epsilon]Y_n(t)\lambda_0(t)\exp[\theta' h]\,dt\xrightarrow{p}\hat{0}$$

for all ϵ. This follows from assumption 4. *QED.*

The essence of the proof is that for the purposes of establishing consistency and asymptotic normality the partial likelihood function can be treated as a conventional likelihood function. Under the assumptions made before, its limit has a unique maximum at θ^*. Also, the normalized derivative

$$N^{-1/2}(\partial\ln\mathscr{L}_p/\partial\theta)(\theta^*)$$

has in the limit a standard normal distribution. A Taylor expansion of the derivative of the log of the partial likelihood function can then be used to obtain the asymptotic distribution for $\hat{\theta}$.

B. Computational Aspects of Partial Likelihood Estimation

A typical computer program for estimation of a Cox regression model with time-varying regressors requires as input the triple $(T, D, X(t), 0<t<T)$, with T the duration, D the censoring indicator, and X the time path of the regressors. The partial likelihood function for this model is, if we order the observations by duration:

$$\mathscr{L}_p(\theta)=\prod_{n=1}^{N}\left[\frac{\exp(\theta' X_n(t_n))}{\sum_{m\geq n}\exp(\theta' X_m(T_n))}\right]^{D_n}. \tag{23}$$

The calendar time dependence model we are interested in consists of observation of the form (t^0, t^1, d, x), with t^0 the time of entry, t^1 the time of exit, d the censoring indicator, and x the regressors. The partial likelihood function corresponding to this model is, ordering the observations by exit time t^1:

$$\mathscr{L}_p(\gamma)=\prod_{n=1}^{N}\left[\frac{\exp(\gamma' h(t_n^1-t_n^0,x_n))}{\sum_{m\geq n,\,t_m^0<t_n^1}\exp(\gamma' h(t_n^1-t_m^0,x_m))}\right]^{d_n}. \tag{24}$$

To estimate this as a standard Cox regression model choose a time origin T_0 such that $t_n^1>T_0$ for all observations. Then let $T=t^1-T_0$ be the (artificial) duration. The censoring indicator D is equal to d. The vector of time varying regressors $X(t)$ consists of two parts. The first is equal to $h(t-t^0,x)$ for $t>t^0$ and equal to zero for $t\leq t^0$. The second part consists of one element and is equal to 1 if $t\leq t^0$ and 0 if $t>t^0$. The partial likelihood corresponding to this specification is, with the observations ordered by (artificial) duration T, which is the same as the ordering by exit time t^1,

$$\mathscr{L}_p(\theta)=\prod_{n=1}^{N}\left[\frac{\exp(\theta' X_n(T_n))}{\sum_{m\geq n}\exp(\theta' X_m(T_n))}\right]^{D_n}$$

$$\mathscr{L}_p(\theta)=\prod_{n=1}^{N}\left[\frac{\exp(\theta_1' h(t_n^1-t_n^0,x_n)+\theta_2\cdot 0)}{\sum_{m\geq n}\exp\left(l_{t_m^0<t_n^1}[\theta_1' h(t_n^1-t_m^0,x_m)+\theta_2\cdot 0]+l_{t_m^0\geq t_n^1}[\theta_1\cdot 0+\theta_2]\right)}\right]^{d_n}.$$

The derivative of the partial likelihood function with respect to θ_2 is always negative, so the coefficient will converge to $-\infty$. When θ_2 gets close to $-\infty$, the partial likelihood function converges to

$$\mathcal{L}_p(\theta) = \prod_{n=1}^{N} \left[\frac{\exp(\theta_1' h(t_n^1 - t_n^0, x_n))}{\sum_{m \geq n,\, t_m^0 < t_n^1} \exp(\theta_1' h(t_n^1 - t_m^0, x_m))} \right]^{d_n}$$

which is the same as (24). The maximand $\hat{\theta}_1$ is therefore identical to the maximand $\hat{\gamma}$ of (24).

REFERENCES

Andersen, P. K., and Richard Gill, "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics* 10 (1982), 1100–1120.

Borjas, G., and G. Sueyoshi, "A Two-stage Estimator for Probit Models with Structural Group Effects," mimeo, University of California at San Diego, 1991.

Breslow, N., "Contribution to the Discussion of the Paper by D. R. Cox," *Journal of the Royal Statistical Society*, Series B, 34 (1972), 187–220.

______, "Covariance Analysis of Censored Survival Data," *Biometrics* 30 (1974), 89–99.

Burdett, Ken, Nick Kiefer, Dale Mortensen, and George Neumann, "Earnings, Unemployment, and the Allocation of Time Over Time," *Review of Economic Studies* 51 (1984), 559–578.

Cox, David R., "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, series B 34 (1972), 187–220.

______, "Partial Likelihood," *Biometrika* 62 (1975), 269–276.

Cox, David R., and David Oakes, *Survival Analysis* (London: Chapman and Hall, 1984).

Elbers, Chris, and Geert Ridder, "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies* 49 (1984), 403–409.

Flinn, Chris, and James J. Heckman, "Models for the Analysis of Labor Force Dynamics," *Advances in Econometrics* 1 (1982), 35–95.

______, "Are Unemployment and Out of the Labor Force Behaviorally Distinct States?" *Journal of Labor Economics* 1 (1) (1983), 28–42.

Gill, Richard D., "Censoring and Stochastic Integrals," *Mathematical Centre Tracts 124*, Mathematisch Centrum Amsterdam, 1980.

Heckman, James J., and Burton Singer, "Econometric Analysis of Longitudinal Data," in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics* 3 (Amsterdam: North-Holland, 1986).

Honoré, Bo, "Simple Estimation of a Duration Model with Unobserved Heterogeneity," *Econometrica* 58 (1990).

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy* 87 (1979), 972–990.

Kay, R., "The Analysis of Transition Times in Multistate Stochastic Processes Using Proportional Hazard Regression Models," *Commun. Statist.-Theor. Meth.*, 11 (15) (1982), 1743–1756.

Lancaster, Tony, "Econometric Methods for the Duration of Unemployment," *Econometrica* 47 (1979).

______, *The Econometric Analysis of Transition Data* (Cambridge: Cambridge University Press, 1990).

Lynch, Lisa, "The Youth Labor Market in the Eighties: Determinants of Re-employment Probabilities for Young Men and Women," this REVIEW 71 (1989), 37–45.

Meyer, Bruce, "Semiparametric Hazard Estimation," mimeo, MIT, 1986.

Miller, Robert A., "Job Matching and Occupational Choice," *Journal of Political Economy* 92 (1984), 1086–1120.

Oakes, David, "Survival Times: Aspects of Partial Likelihood," with discussion, *International Statistical Review* 49 (1981), 235–264.

Ridder, Geert, *Life Cycle Patterns in Labor Market Experience*, unpublished PhD dissertation, University of Amsterdam, 1987.

Self, S. G., and R. L. Prentice, "Commentary on Andersen and Gill's 'Cox's Regression Model for Counting Processes: A Large Sample Study'," *The Annals of Statistics* 10 (1982), 1121–1124.

Tsiatis, A. A., "A Large Sample Study of Cox's Regression Model," *The Annals of Statistics* 9 (1981), 93–108.

Van Den Berg, Gerard, "Nonstationary in Job Search Theory," *Review of Economic Studies* 57 (2) (1990).