Consistent Estimators for Panel Duration Data with Endogenous Censoring and Endogenous Regressors[¤]

Tiemen M. Woutersen
Brown University

January 3, 2000

Abstract

The estimation of duration models is complicated by unobserved heterogeneity and right censoring of the data. Assumptions on the unobserved heterogeneity can be avoided by using a ¯xed e®ect parametrization. This paper develops new estimators that are consistent under right censoring for a wide range of parametric duration models and also derives a nonparametric estimator for the baseline hazard that converges in mean square error. Moreover, the new estimator allows for endogenous regressors and enables to distinguish state dependence from heterogeneity.

1 Introduction

The analysis of the durations of events has been a subject of some attention in econometrics for almost two decades. It has been complicated by heterogeneity of the data and by the fact that economic agents choose regressors given their characteristics and history. Thus the regressors depend on the heterogeneity and may be endogenous as well. A further complication is that a part of the data may be censored. This paper develops estimators for this framework.

The problem of heterogeneity is often dealt with by making distributional assumptions on the unobserved heterogeneity. One of the assumptions usually is that the unobserved heterogeneity and regressors are independent. In that case, estimation by maximum likelihood is possible. Lancaster (1990) gives an overview. However, both Heckman and Singer (1984) and Lancaster (1990) have demonstrated that the estimated parameters can be sensitive to the assumed functional form of the heterogeneity. We can avoid these functional form assumptions by using a ¯xed e®ect speci¯cation: a time-invariant, individual parameter for each individual. Honoré (1993) gives identi¯cation results for duration models with multiple, uncensored spells. Lancaster (1999) suggests ¯nding a parametrization of the ¯xed e®ects

[¤] I am very grateful to Tony Lancaster, Moshe Buchinsky, Andrew Foster and Guido Imbens for helpful comments and discussions. All errors are mine.

that is orthogonal in the information matrix sense and then integrating out these ¯xed effects. When we apply Lancaster's method to the ¯xed e®ect Weibull model with uncensored spells, we get Chamberlain's (1985) consistent estimator for the Weibull model.

However, Lancaster's method requires a di®erential equation to be solved analytically. This is not possible for the Weibull or exponential duration model in the case that we have right censoring. In general, right censoring complicates the analysis of panel duration data. In particular, the chance that a duration spell is censored depends on the length of all earlier spells and is therefore not exogenous. Indeed, Van den Berg (1999) notes that there are hardly any estimation methods for panel duration data with censoring, whether or not we allow for a ¯xed e®ect. We want to allow for ¯xed e®ects and right censoring. This paper develops estimators for all important duration models. That is, for the exponential, Weibull, piecewise-constant and log-linear hazard models as well as the nonparametric baseline hazard with exponential regressors. We allow for interaction between the di®erent spells of the same individual in order to distinguish between state dependence and heterogeneity.

In the case that we only have one spell for each individual, we cannot allow for an individual parameter for each individual since none of our parameters of interest would be identi¯ed. Depending on the application, we may be willing to assume that the value of the ¯xed e®ect depends on certain observable characteristics of the individual. If the data allow us to ¯nd other individuals with similar characteristics, then we can construct a new `panel' with the duration of the individual and a `weighted average' of other individuals with similar characteristics. Obviously, the interesting interaction between spells of the same individual is lost for the single spell case, but consistent estimation of a number of parametric models is still possible. To be more precise: for the parametric models without interaction between spells we can substitute the multiple spell requirement with a semiparametric matching condition.

Most of the estimators derived in this paper deal with cases where, so far, no consistent estimator existed. The exception is a rank estimator for the slope coe±cients of the proportional hazard model as derived by Chamberlain (1985). Ridder and Tunali (1999) use this estimator for censored data. Crucial assumptions for this estimator are that the censoring is exogenous and that there is no interaction between spells. Within this framework we compare the asymptotic variance of this rank estimator to the asymptotic variance of the estimators derived in this paper.

The paper is organized as follows. Section 2 shows how inference can be based on the integrated hazard. It develops theorems for consistent estimators for several duration models with ¯xed e®ects and right censoring. The nonparametric estimation of the baseline hazard in this section and all theorems of further sections are based on these theorems. Section 3 then deals with the cases in which the chance of right censoring depends on the length of previous spells. Section 4 deals with endogenous regressors like functions of previous unemployment and employment spells. Section 5 substitutes the multiple spell requirement with a matching condition, so only one spell is needed. Section 6 compares estimators and section 7 concludes.

2 Estimation using the integrated hazard

In this section we use properties of the integrated hazard function to derive estimators. First we give two easy examples to illustrate the method (section 2.1), then we give theorems for several parametric hazard specifications (section 2.2) and finally we explain how to use the same method for nonparametric estimation of the hazard function (section 2.3).

2.1 Easy examples

This subsection is only meant to illustrate how to derive estimators that are based on the integrated hazard; the resulting estimators are a very common maximum likelihood and a methods of moments estimator. In the next subsections we use the integrated hazard to derive estimators for cases where, so far, no estimator existed.

We frequently use the fact that, given the exogenous regressor x; the integrated hazard is a unit Exponential variate.

$$Z = \int_0^t \theta(s; x) ds \sim \mathcal{E}(1). \tag{1}$$

So the expectation of Z equals one.

Example 1. We can use equation (1) to estimate parameters of the hazard function. Consider almost the simplest problem to illustrate how this is done. Assume that $t_1; \ldots; t_N$ are independent durations with hazard $\theta(t) = e^{\beta}$ so $z = e^{\beta} t$. The integrated hazards are independent unit exponentials, $z = e^{\beta} t \sim \mathcal{E}(1)$: We estimate the model by choosing a β such that the integrated hazard has expectation one. Equating the sample analogue of the integrated hazard to one gives:

$$\frac{\sum_{i=1}^{N} e^{\beta} t_i}{N} = 1.$$

This gives an estimate for β;

$$\hat{\beta} = \ln\left(\frac{\sum_{i=1}^{N} t_i}{N}\right);$$

which is the maximum likelihood estimator.

Example 2. Suppose that we observe two spells for N individuals and want to estimate an exponential hazard model. Assume that we observe for each period a vector of regressors for individual i: Since we have more than one observation for each individual we can allow for a fixed effect. In the fixed effect parametrization, the coefficients of the regressors depend on the difference of the regressors in the first and second spell. Let the vector of these differences be denoted by Δx_i and let ΔX be a matrix with the vectors Δx_i; $i = 1; \ldots; N$; as its rows. Assume ΔX has full column rank. If the spells are independent across individuals as well as across spells for the same individual then the hazard of the first and second spell can be written as follows:

$$\begin{aligned} \theta_{i1} &= f_i \\ \theta_{i2} &= f_i e^{\Delta x_i \beta} . \end{aligned}$$

The integrated hazards for the first and second spell are $\phi_i t_{i1}$ and $\phi_i e^{\Delta x_i \beta} t_{i2}$, respectively. At the true parameter value, β_0, the difference between these integrated hazards equals zero in expectation. The expectation of the difference does not depend on the value of the fixed effect and therefore

$$E(t_{i1} - e^{\Delta x_i \beta_0} t_{i2}) = 0.$$

Multiplying by the vector Δx_i gives

$$\Delta x_i E(t_{i1} - e^{\Delta x_i \beta_0} t_{i2}) = 0.$$

The last equation suggests the following moment vector function

$$g(\beta) = \frac{1}{N} \sum_i g_i$$

where

$$g_i(\beta) = \Delta x_i (t_{i1} - e^{\Delta x_i \beta} t_{i2}).$$

Estimation based on the moment vector function $g(\beta)$ gives consistent estimates for β when the number of individuals goes to infinity[1]. Similar moment functions can be derived by taking the logarithms of the durations and then differencing out the fixed effect. Estimators that are based on the log durations need to observe the actual duration. An advantage of inference based on integrated hazards is that it can deal with censoring, i.e. that resulting estimators are consistent. Another advantage of the integrated hazard is that it suggests new consistent estimators for studying the interactions between spells. We show these advantages in the remaining part of this section and in the next section.

2.2 Estimation of several parametric hazard specifications

In this subsection we use the integrated hazard to derive estimators for several parametric hazard models. Since data are often right censored and heterogenous we allow for fixed effects and right censoring. Van den Berg mentions in the forthcoming Handbook of Econometrics an "important caveat" with multiple spell data: Existing estimator are "particularly sensitive to censoring". Indeed, even for the Weibull, or piecewise-constant hazard model a consistent estimator was lacking. This section derives consistent estimators for these and other models.

Suppose we observe the minimum of the duration t_{is} and the censoring time c_{is}. Let us denote this observed length of a spell by y_{is}. In this subsection we assume that the censoring time is exogenous, i.e. the censoring time does not depend on the length of the current or past spells. In section 3 we will discuss estimators with endogenous censoring.
For now, y_{is} is determined as

$$y_{is} = \min(t_{is}, c_{is}).$$

[1] A proof of a more general case is given in section 2.2.1.

The indicator d_{is} equals zero when the observation is censored and equals one when the observation is uncensored. The expectation of this indicator, $E(d_{is})$; equals the probability that the observation is not censored. A useful fact is that the expectation of d_{is} equals the expectation of the integrated hazard.

Lemma 1 The expectation of the integrated hazard of a spell that may be right censored equals the expectation of the indicator that denotes no censoring:

$$E(z_{is}) = E \int_0^{y_{is}} µ(s; x)ds = E(d_{is}): \tag{2}$$

Proof: See appendix 1.

2.2.1 Exponential hazard model

Suppose we want to estimate the exponential hazard model and observe two possibly censored spells for N individuals. More precisely, suppose the censoring time is exogenous and denoted by c_{is}; this censoring time may vary over individuals and spells. We observe, for each individual, $y_{is} = \min(t_{is}; c_{is})$; $s = 1; 2$: Assume that we observe exogenous regressors as well. As in example 2, section 2.1, we denote the di®erence between the regressors of the ¯rst and second spell by $4 x_i$ and de¯ne $¢X$ as the matrix with the vectors $4 x_i$; $i = 1; :::; N$; as its rows. Assume that $¢X$ has full column rank. We allow for a ¯xed e®ect f_i and write the hazard of the ¯rst and second spell as follows:

$$\begin{aligned} µ_{i1} &= f_i \\ µ_{i2} &= f_i e^{4 x_i ¯} . \end{aligned}$$

The integrated hazards for the ¯rst and second spell are $f_i y_{i1}$ and $f_i e_{i2}^{4 x_i ¯} y_{i2}$; respectively. The expectations of the integrated hazards are given by lemma 1:

$$\begin{aligned} Ez_{i1} &= E f_i y_{i1} = Ed_{i1} \\ Ez_{i2} &= E f_i e^{4 x_i ¯_0} y_{i2} = Ed_{i2} \\ \text{in general, } Ed_{i1} &\neq Ed_{i2}: \end{aligned}$$

In general, the probability of a spell being censored di®ers for the ¯rst and second period[2]. By lemma 1, this implies that the expectation of the integrated hazard of the ¯rst and second period di®ers as well. For uncensored spells, the expectation of the integrated hazard was the same for both periods (and equal to one). In section 2.1, example 2, we used this fact to derive an estimator for the exponential hazard model. For the censored case, however, we need to adjust the method to take care for the inequality of the expected integrated hazards. Basically, the integrated hazard with the largest expectation needs to be decreased to restore equality of the expectations. This can be done by multiplying the integrated hazard of the

[2] If $c_{i1} = c_{i2}$ and $4 x_i = 0$ then we know a priori that the probability of censoring is the same for the ¯rst and second spell. However, only individuals with $4 x_i \neq 0$ are useful for the estimation of ¯:

first period by d_{i2} and multiplying the integrated hazard of the second period by d_{i1}. The expression that results from this multiplication has the following expectation:

$$\begin{aligned} E(d_{i2}z_{i1} - d_{i1}z_{i2}) &= E(d_{i2}z_{i1}) - E(d_{i1}z_{i2}) \\ &= Ed_{i2}Ez_{i1} - Ed_{i1}Ez_{i2} \end{aligned}$$

where the last line follows from the independence of the first and second spell. From lemma 1 it follows that $Ez_{i1} = Ed_{i1}$ and $Ed_{i2} = Ez_{i2}$. Therefore

$$E(d_{i2}z_{i1} - d_{i1}z_{i2}) = 0. \tag{3}$$

In example 1 and example 2 we based an estimator on the difference of the integrated hazards of the first and second spell[3]. That difference was zero in expectation and the resulting estimators were consistent. Equation (3) suggests an adjustment for exogenous censored spells. Indeed, estimators based on (3) are in general consistent. The remainder of this section is devoted to show this consistency for several duration models. For each model we write the integrated hazard in the form that is implied by the model. For the exponential model rewriting equation (3) gives:

$$E(d_{i2} f_i y_{i1} - d_{i1} f_i e^{\Delta x_i \beta_0} y_{i2}) = 0.$$

The expectation does not depend on the value of the fixed effect f_i and therefore

$$E(d_{i2} y_{i1} - d_{i1} e^{\Delta x_i \beta_0} y_{i2}) = 0.$$

As in example 2, we multiply by Δx_i and get

$$\Delta x_i E(d_{i2} y_{i1} - d_{i1} e^{\Delta x_i \beta_0} y_{i2}) = 0.$$

The last equation suggests the following vector moment function:

$$g(\beta) = \frac{1}{N} \sum_i g_i \tag{4}$$

where

$$g_i(\beta) = \Delta x_i (d_{i2} y_{i1} - d_{i1} e^{\Delta x_i \beta_0} y_{i2}). \tag{5}$$

We define the objective function $Q(\beta) = g(\beta)' g(\beta)$. Maximizing $Q(\beta)$ w.r.t. β gives a consistent estimate for β. We will proof this in theorem 1. Theorem 1 will also cover the cases in which the number of spells per individual differs from two. For example, suppose we observe three possibly censored spells for each individual. Let x_{is} denote the exogenous vector of regressors of spell s of individual i. We can use the moment function of equation (5) and use only the first two spells. That is, we use the following moment function

$$g_i^A(\beta) = (x_{i2} - x_{i1})(d_{i2} e^{x_{i1}\beta} y_{i1} - d_{i1} e^{x_{i2}\beta} y_{i2}).$$

[3] See 2.1 Easy examples.

Similarly, we could choose two other pairs of spells:

$$\begin{aligned} g_i^B(\beta) &= (x_{i3} - x_{i2})(d_{i3}e^{x_{i2}\beta}y_{i2} - d_{i2}e^{x_{i3}\beta}y_{i3}) \\ g_i^C(\beta) &= (x_{i3} - x_{i1})(d_{i3}e^{x_{i1}\beta}y_{i1} - d_{i1}e^{x_{i3}\beta}y_{i3}). \end{aligned}$$

A moment function that uses all spells is, obviously,

$$g_i(\beta) = g_i^A(\beta) + g_i^B(\beta) + g_i^C(\beta).$$

Let the number of spells of individual i be denoted by T_i. The general expression for $g_i(\beta)$ allows for a different number of spells per individual:

$$g_i(\beta) = \sum_{s<r} \{(x_{ir} - x_{is})(d_{ir}e^{x_{is}\beta}y_{is} - d_{is}e^{x_{ir}\beta}y_{ir})\} \quad s, r = 1, \ldots, T_i.$$

Inference is based on the moment function $g(\beta) = \frac{1}{N}\sum_i g_i$. At the true value of the parameter, β_0, this moment function has expectation zero:

$$\begin{aligned} Eg_i(\beta_0) &= E\sum_{s<r}\{(x_{ir} - x_{is})(d_{ir}e^{x_{is}\beta_0}y_{is} - d_{is}e^{x_{ir}\beta_0}y_{ir})\} \\ &= \frac{1}{f_i}\sum_{s<r}\{(x_{ir} - x_{is})(Ed_{ir}Ez_{is} - Ed_{is}Ez_{ir})\} \text{ using independence} \\ &= 0 \text{ since } Ed_{ir} = Ez_{ir} \text{ and } Ed_{is} = Ez_{is}. \end{aligned}$$

Therefore

$$Eg(\beta_0) = \frac{1}{N}\sum_i Eg_i(\beta_0) = 0.$$

Panel duration data typically have a large number of individuals observed over a short time period. Therefore, the relevant limiting distributions have the number of individuals, N, increasing though not the time dimension[4]. So the number of spells is fixed but may vary over individuals. Because the number of spells may vary over individuals this makes the 'full rank' condition slightly more difficult than in the case where we just had two spells. We define X as a matrix with $\sum_i T_i$ rows and K columns. The first row contains the regressors of the first spell of the first person, i.e. x_{11}', the second row contains the regressors of the second spell of the first person,..., the $(T_1 + 1)^{th}$ contains the regressors of the first spell of the second person and so on. That is

$$X = \begin{pmatrix} x_{11}' \\ x_{12}' \\ \vdots \\ x_{NT_N}' \end{pmatrix}.$$

[4] See Chamberlain 1985.

When we assume that X has full column rank then we are able to estimate β consistently.

Theorem 1
Assume that we observe at least two spells for N individuals, that all the spells are independent given the regressors and have the following hazards

$$\theta_{is}(t) = f_i e^{x_{is}\beta} \quad s = 1, \ldots, T_i.$$

The censoring time, c_{is}, is exogenous and we observe $y_{is} = \min(t_{is}, c_{is})$. Further assume that $\beta \in \Theta$, which is compact, and that X has full rank. Define $Q(\beta) = -g(\beta)'g(\beta)$. Maximizing $Q(\beta)$ w.r.t. β gives a consistent estimate for β, i.e.

$$\hat{\beta} = \arg\max_{\beta} Q(\beta) \to_p \beta_0 \text{ for } N \to \infty.$$

Proof: See appendix 2.

2.2.2 General theorem for one-parameter hazards

By using the insights of the exponential hazard model we can develop a general theorem for single parameter hazard rates. Let the hazard depend on the scalar parameter γ. Assume the spells are independent spells and the censoring times are exogenous. The following lemma holds for all one parameter duration models.

Lemma 2
Define $g(\gamma) = \frac{1}{N f_i} \sum_{i=1}^{N} \{ \sum_{s<r} (d_{ir} z_{is} - d_{is} z_{ir}) \}$ where $z_{is} = \int_0^{y_{is}} \theta(s; x) ds$. Assume that the spells are independent across both individuals and spells, and that the censoring time is exogenous. Then

$$E g(\gamma_0) = 0.$$

Proof: See appendix 3.

Lemma 2 only provides one moment function. However, the econometrician can censor the data artificially and apply lemma 2 several times. We show in the remaining subsections how this can be done; in section 2.3 we develop an estimator based on lemma 2 which estimates the baseline hazard nonparametrically. The basis for more advanced estimators is lemma 2 and the following theorem. This theorem states the conditions under which the moment function of lemma 2 gives consistent estimates of γ, the parameter of interest.

Theorem 2
Define $Q_n(\gamma) = -g(\gamma)^2$. If $g(\gamma)$ is continuous in γ and if either assumption I or II is satisfied,

I. $g(\gamma)$ is monotonic and $E \sup_{\gamma} ||g(\gamma)|| < \infty$;

II. $\gamma \in \Theta$, which is compact; $E g(\gamma) = 0$ only if $\gamma = \gamma_0$ and $E \sup_{\gamma \in \Theta} ||g(\gamma)|| < \infty$;

then $\hat{\gamma} = \arg\max_{\gamma} Q_n(\gamma)$ is a consistent estimate for γ_0, i.e.

$$\hat{\gamma} = \arg\max_{\gamma} Q_n(\gamma) \to_p \gamma_0.$$

Proof:
$Eg(\gamma_0) = 0$ by lemma 2. If assumption I is satisfied then $g(\gamma)$ is monotonic and $Q(\gamma) = -g(\gamma)^2$ is concave in γ and therefore $EQ(\gamma)$ is uniquely maximized at the truth. Therefore, maximizing $Q_n(\gamma)$ gives a consistent estimate; see Newey and McFadden, 1994, Theorem 2.7. If assumption II is satisfied then the conditions of Newey and McFadden, 1994, Theorem 2.6 are satisfied and maximizing $Q_n(\gamma)$ gives a consistent estimate of γ. Q.E.D.

We are not aware of hazard rates that are not continuous in their parameters. In the next subsections, we check for several hazard models in which either condition I or condition II is satisfied.

2.2.3 Piecewise-constant hazard

In many countries, unemployed people can only collect unemployment benefits for a certain amount of time. To test whether the hazard rate depends on being eligible for unemployment benefits, we can estimate a piecewise-constant hazard model. Assume that the government provides unemployment benefits for u_1 periods and that we observe two spells for N individuals. A model to estimate the effect of unemployment benefits on the hazard is[5]

$$\begin{aligned} \theta(t_{is}) &= f_i \quad \text{if } t_{is} \leq u_1 \\ &= f_i\gamma \quad \text{if } t_{is} > u_1. \end{aligned}$$

The integrated hazard equals

$$\begin{aligned} z_{is} &= \int_0^{t_{is}} f_i ds = f_i t_{is} \text{ if } t_{is} \leq u_1 \\ &= f_i u_1 + f_i\gamma(t_{is} - u_1) \text{ if } t_{is} > u_1. \end{aligned}$$

The econometrician can censor the second spell at u_1; the number of periods for which the government provides unemployment benefits. This ensures monotonicity of $g(\gamma)$:

$$g(\gamma) = \frac{\sum_i}{N}\left\{\frac{d_{i2}z_{i1} - d_{i1}z_{i2}}{f_i}\right\} \tag{6}$$

$$\frac{\partial g(\gamma)}{\partial\gamma} = -\frac{\sum_i}{N}\frac{d_{i2}}{f_i}\frac{\partial z_{i1}}{\partial\gamma} = -\frac{\sum_i}{N} d_{i2}\, m(t_{i1} > u_1)$$

where $m(t_{i1} > u_1)$ is an indicator that equals one if $\{t_{i1} > u_1\}$ and zero otherwise. So $\frac{\partial g(\gamma)}{\partial\gamma} < 0$ if there is at least one t_{i1} that is larger than u_1. Consistency follows from theorem 2.

[5] Exogenous regressors can be added to this model in the same way as was done for the Weibull model in section 2.2.4.

The estimation of a piecewise-constant hazard with more than two `pieces' is similar, e.g. the model with the following hazard

$$\begin{aligned} \theta(t_{is}) &= f_i \quad \text{if } t_{is} \leq u_1 \\ &= f_i \gamma_1 \quad \text{if } u_1 < t_{is} \leq u_2 \\ &= f_i \gamma_2 \quad \text{if } t_{is} > u_2 \end{aligned}$$

is estimated with a repeated use of equation (6). First we censor the first spell at u_2 and the second spell at u_1 and obtain a moment equation that only involves the parameter γ_1. We then censor the second spell at u_2 while leaving the first spell uncensored and the resulting moment function involves both parameters. In subsection 2.3 we discuss this procedure in further detail and let the number of intervals increase and to estimate the baseline hazard nonparametrically.

2.2.4 Weibull hazard

Suppose we observe two possibly censored spells for N observations and that we want to estimate a Weibull model. First we show how this can be done for the Weibull model without regressors; then we discuss the Weibull model with regressors. The hazard has the form $\theta_{is} = \alpha f_i t_{is}^{\alpha-1}$ $(s = 1, 2)$ and the integrated hazard is $z_{is} = f_i t_{is}^{\alpha}$ $(s = 1, 2)$. We observe $y_{is} = \min(t_{is}, c_{is})$. Theorem 2 suggests the following moment function

$$\begin{aligned} g(\gamma) &= \frac{1}{N f_i} \sum_{i=1}^{N} \{ d_{i2} z_{i1} - d_{i1} z_{i2} \} \\ &= \frac{1}{N} \sum_{i=1}^{N} \{ d_{i2} y_{i1}^{\gamma} - d_{i1} y_{i2}^{\gamma} \} \end{aligned} \tag{7}$$

$$\frac{\partial g(\gamma)}{\partial \gamma} = \frac{1}{N} \sum_{i=1}^{N} \{ d_{i2} y_{i1}^{\gamma} \ln y_{i1} - d_{i1} y_{i2}^{\gamma} \ln y_{i2} \}.$$

The equation $Eg(\gamma) = 0$ has a unique solution if $g(\gamma)$ is monotonic. Monotonicity of $g(\gamma)$ is assured if only the second period is censored. The function $g(\gamma)$ is in general monotonic if the probability of the censoring differs for the first and second spell. Obviously, plotting $g(\gamma)$ indicates whether the function is monotonic. If $g(\gamma)$ is relatively flat, as is the case when the probability of censoring is the same for both the first and second spell, then the econometrician can artificially censor the second spell. This means that the econometrician censors the observed data at a censoring points he prefers: Instead of using the observed y_{is}'s the econometrician censors the second period[6] and uses $y'_{i2} = \min(y'_{i2}, \tilde{c}_{i2})$ as data where $\tilde{c}_{i2} \leq c_{i2}$ for at least some i. This artificial censoring assures that $g(\gamma)$ is monotonic and therefore $Eg(\gamma) = 0$ is uniquely solved for $\gamma = \gamma_0$. Furthermore, $E \sup_{\gamma \in \Gamma} \|g(\gamma)\| < 0$ and therefore, condition II of theorem 2 is satisfied and consistency follows.

[6] In general, the spell is choosen which has the largest number of censored observations.

The standard error of the estimator can sometimes be reduced by artificial censoring. Assume that the second spell is censored at the common censoring time c_2. The asymptotic variance of the estimator is then given by $\frac{E[g(\gamma)^2]}{\{E[g_\gamma(\gamma_0)]\}^2}$. The econometrician can artificially censor the data at $\tilde{c}_2$ where $\tilde{c}_2 < c_2$ and redo his estimation including a calculation of the variance of the estimator. A more systematic way is to choose c_2 as the starting value of $\tilde{c}_2$ and then let $\tilde{c}_2$ decrease in small steps. For every step an estimate for γ is calculated by using the moment function of equation (7); the asymptotic variance of the estimator is calculated as well. A sensible procedure is to lower $\tilde{c}_2$ until the asymptotic variance, $\frac{E[g(\gamma)^2]}{\{E[g_\gamma(\gamma_0)]\}^2}$, does not decrease anymore.

We can allow for regressors by combining the moment function of equation (7) with the moment function of theorem 1. We stack those moment functions and define $g^*(\beta;\gamma)$:

$$g^*(\beta;\gamma) = \begin{pmatrix} g_W(\beta;\gamma) \\ g_E(\beta;\gamma) \end{pmatrix}$$

where

$$g_W(\beta;\gamma) = \frac{1}{N}\sum_{i=1}^{N}\{\sum_{s<r}(\tilde{d}_{ir}e^{x_{is}\beta}\tilde{y}_{is}^{\gamma} - \tilde{d}_{is}e^{x_{ir}\beta}\tilde{y}_{ir}^{\gamma})\}$$

and

$$g_E(\beta;\gamma) = \frac{1}{N}\sum_{i=1}^{N}[\sum_{s<r}\{(x_{ir} - x_{is})(d_{ir}e^{x_{is}\beta}y_{is}^{\gamma} - d_{is}e^{x_{ir}\beta}y_{ir}^{\gamma})\}].$$

The superscript for d_{is} and $\tilde{y}_{is}$ denotes that the data are possibly artificially censored. The function $g_E(\beta;\gamma)$ uses the data in their original form, i.e. no artificial censoring takes place for that part of the moment vector function.

The moment vector function $g^*(\beta;\gamma)$ has expectation zero at the truth, $\{\beta_0;\gamma_0\}$ and the minimizing rest $Q(\beta;\gamma) = -\{g^*(\beta;\gamma)\}'\{g^*(\beta;\gamma)\}$ gives a consistent estimate of the parameters (see appendix 4).

For the Weibull model without regressors there is another estimator available, so we will compare the assumptions and properties of the different estimators. Honoré (1990) develops an estimator for the Weibull parameter that is based on an order statistic. He uses a random effects model and the only assumption about the mixing distribution is that its mean is finite. Only one spell per person is needed. Asymptotically, the method uses only observations close to zero; so theoretically it is not hindered by censoring. However, in some applications, e.g. in demography Heckman, Hotz and Walker (1985) in which the amount of time between first and second birth, etc. is studied, one does not want to rely too much on observations close to zero since one does not know when "time zero" happens to be. In this case, fertility after giving birth is unobservable and lags behind the observable event of birth in a person specific way. With unemployment data time zero can be the moment one starts looking for a new job; related events could be the moment one loses a job or gets unemployment benefits. Thus the disadvantages are that we let the shape parameter be determined by the shape in a small interval $[0;\epsilon]$ and that we sometimes do not know its location. Moreover, Van den

Berg (1999) notes that extremely short durations are often under-reported in real-life data and that it is "notoriously difficult" to assess the shape of the hazard function for $t \downarrow 0$; let alone the extrapolation. However, the importance of these disadvantages depends on the application.

As the following theorem shows, the rate of convergence of our method is $\sqrt{N}$. In contrast, the maximum rate of convergence that the method of Honoré can attain is $N^{1/3}$; which implies that there is an advantage in the use of panel data in both robustness and rate of convergence.

Theorem 3
Define $Q_n(\gamma) = -g(\gamma)^2$. Assume $\gamma \in \Theta$; which is compact. Then the estimator $\hat{\gamma} = \arg\max Q_n(\gamma)$ converges at rate $\sqrt{N}$ to the true value γ_0. Moreover, $\sqrt{N}(\hat{\gamma} - \gamma_0) \to N(0; \frac{E[g(\gamma)^2]}{\{E[g_\gamma(\gamma_0)]\}^2})$.
Proof:

(i) $g(\gamma)$ is continuously differentiable in γ;
(ii) $E(g(\gamma_0)) = 0$ and $E(|g(\gamma_0)|^2) < \infty$;
(iii) $E[\sup_{\gamma \in \Theta} |g_\gamma(\gamma)|] < \infty$;
(iv) $\{E[g_\gamma(\gamma_0)]\}^2$ is nonsingular.

These are the standard conditions for asymptotic normality of the GMM.[7] Q.E.D.

2.2.5 Log-linear hazard and the unemployment rate as regressor

The logarithm of the hazard is linear in the parameter γ:

$$\ln \theta(x;t) = \ln(f_i) + \gamma k(x(t);t)$$

where $k(x(t);t)$ is a function of the regressors and the data. Assume that we observe two spells for N individuals. Lemma 2 suggests to use the following moment function

$$\begin{aligned} g(\gamma) &= \frac{\sum_i \{\frac{d_{i2} z_{i1} - d_{i1} z_{i2}}{f_i}\}}{N} \\ &= \frac{\sum_i \{d_{i2} \int_0^{y_{i1}} \exp(\gamma k(x(s);s))ds - d_{i1} \int_0^{y_{i2}} \exp(\gamma k(x(s);s))\}}{N}; \text{ and} \end{aligned} \tag{8}$$

$$\frac{\partial g(\gamma)}{\partial \gamma} = \frac{\sum_i \{d_{i2} \int_0^{y_{i1}} k(x(s);s) \exp(\gamma k(x(s);s))ds - d_{i1} \int_0^{y_{i2}} k(x(s);s)) \exp(\gamma k(x(s);s))\}}{N}.$$

Plotting the function $g(\gamma)$ reveals whether the solution to $g(\gamma) = 0$ is unique. If it is not we can try to artificially censor the data and plot $g(\gamma)$ again. Assuming that $Eg(\gamma) = 0$ is uniquely solved for $\gamma = \gamma_0$ and noting that $E \sup_{\gamma \in \Theta} ||g(\gamma)|| < 0$; then theorem 2 assures consistency.

In the case that the second period is hardly censored then $g(\gamma)$ may be relatively flat and $\frac{\partial g(\gamma)}{\partial \gamma}$ is small and, as a consequence, the standard error of γ will be large. This standard

[7] See e.g. Newey and McFadden (1994), page 2148.

error can sometimes be reduced by artificially censoring the second period, i.e. use $\tilde{y}_{i2} = \min(y_{i2}, \tilde{c}_2)$. Without artificial censoring we have $\tilde{c}_2 = c_2$. Analogue to the Weibull case we lower $\tilde{c}_2$ until the asymptotic variance of γ does not decrease anymore. The asymptotic variance of γ equals $\frac{E[g(\gamma)^2]}{\{E[g_\gamma(\gamma_0)]\}^2}$.

A special case of the log-linear hazard model is the exponential model with time-varying regressors. Theorem 1 gave a moment vector function for the exponential model. The value of the exogenous regressors could vary over different spells but was restricted to be constant within a spell. However, some explaining variables, like the unemployment rate, do not stay constant during an unemployment spell of an individual. We could use the moment function of equation (8) to estimate the coefficient of unemployment. From Lemma 2 follows that $Eg(\gamma_0) = 0$. To assure uniqueness of $Eg(\gamma_0) = 0$ we can censor artificially but we can also multiply with functions of the unemployment rate. Note that if a random variable has expectation zero then that random variable still has expectation zero when we multiply it with a function of exogenous variables. A function of exogenous variables is the difference between the average unemployment rate in the first and second spell. Let us denote the average unemployment rate during the s^{th} spell of individual i by u_{is}. Then an adjustment of equation (8) could be

$$g(\gamma) = \frac{\sum_i}{N}(u_{i2} - u_{i1})\left\{\frac{d_{i2}z_{i1} - d_{i1}z_{i2}}{f_i}\right\}.$$

The generalization to T_i spells is straightforward: Just replace x_{is} by u_{is} in the moment function of theorem 1. In case we have several time dependent regressors then we need to censor the data artificially to create as many moments as time dependent variables[8].

2.3 Nonparametric and semi-nonparametric estimation of the baseline hazard

In this subsection we use theorem 2 to estimate the baseline hazard nonparametrically. Some writers have been skeptical about estimating the baseline hazard nonparametrically since that estimation is very hard to interpret as long as we do not take account for the heterogeneity in the data. For example, Lancaster (1990) devotes a section of his book to nonparametric estimation and notes "since data are rarely homogeneous this section is not of direct relevance to econometric work". Indeed, even if we specify the baseline hazard to be the Weibull hazard then it turns out that the estimate of the Weibull coefficient can change when we add regressors that take account of heterogeneity. In Lancaster (1979) the inclusion of more regressors changed a decreasing hazard to a nearly constant one (using a Weibull specification). So estimation of the baseline hazard is sensitive to whether we take the heterogeneity into account. In this section we react to this "heterogeneity-sensitivity" by developing a nonparametric estimator for the baseline hazard that allows for fixed effects. This section is organized as follows. First we extend the example of the piecewise-constant hazard function of the last section by shrinking the length of the intervals over which the hazard is constant. The length of the intervals can be arbitrary short and therefore the

[8] Plus moments for the Weibull parameter and the parameters for the interactions between spells.

resulting estimator can be called nonparametric. The next step is to combine the estimation of the baseline hazard with theorem 1. The resulting theorem allows to estimate the baseline hazard and the coefficients for the regressors simultaneously while allowing for fixed effects and right censoring. Suppose that the durations are independent and that the hazard has the following form

$$\theta(t_{is}) = f_i \lambda_0(t)$$

where $\lambda_0(t)$; the baseline hazard, is an unknown function of t: We observe

$$y_{is} = \min(t_{is}; c)$$

where c is the common, exogenous censoring time. We divide the interval $[0; c]$ into m intervals. The number of intervals, m; increases with the number of individuals; we choose $m = [kn^{1=5}]$ where [] means 'the next integer smaller than the argument' and k denotes a constant. The length of each interval is $r = \frac{c}{m}$: To estimate the function $\lambda_0(t)$ we estimate the following hazard model

$$\begin{aligned}
\theta(t_{is}) &= f_i \quad \text{if } t_{is} \le r \\
&= f_i \gamma_1 \quad \text{if } r < t_{is} \le 2r \\
&= f_i \gamma_2 \quad \text{if } 2r < t_{is} \le 3r \\
&\;\;\vdots \\
&= f_i \gamma_{m-1} \quad \text{if } (m-1)r < t_{is} \le c \text{ where } m = [kn^{1=5}]:
\end{aligned}$$

Note that if $n \to \infty$ then the number of intervals goes to infinity, i.e. $m = [kn^{1=5}] \to \infty$: However, the number of intervals goes to infinity at a slower rate than n so their ratio, $\frac{[kn^{1=5}]}{n} \to 0$ and, loosely put, the number of observations per interval increases. The integrated hazard of this approximation that we estimate is

$$\begin{aligned}
z_{is} &= \int_0^{t_{is}} f_i ds \\
&= f_i t_{is} \text{ if } t_{is} \le r \\
&= f_i(r + (t_{is} - r)\gamma_1 \text{ if } r < t_{is} \le 2r \\
&= f_i(r(1 + \gamma_1) + (t_{is} - 2r)\gamma_2 \text{ if } 2r < t_{is} \le 3r \\
&= f_i(r(1 + \gamma_1 + \gamma_2) + (t_{is} - 3r)\gamma_3 \text{ if } 3r < t_{is} \le 4r \\
&\;\;\vdots \\
&= f_i(r(1 + \sum_{l=1}^{m-1} \gamma_l) + (t_{is} - (m-1)r)\gamma_{m-1} \text{ if } (m-1)r < t_{is} \le c \text{ where } m = [kn^{1=5}] .
\end{aligned}$$

In the last section we artificially censored the second spell at the point where unemployment benefits ended to ensure concavity of the objective function. Thus we had two intervals and artificially censored the data once. In this case we have m intervals and artificially censor the data $m - 1$ times. We use $m - 1$ moment function to estimate all the parameters. For the l^{th} moment function, to estimate γ_l; we artificially censor the data in the following way

$$\begin{aligned}
\text{moment function } l \quad y_{i1} &= y_{i1} \\
y_{i2} &= \min(y_{i2}; l \cdot r):
\end{aligned}$$

Analogue to the last section we have the following estimating equations

$$g_l(\gamma) = \frac{1}{L_n} \sum_i \{\frac{d_{i2}z_{i1} - d_{i1}z_{i2}}{f_i}\} \text{ for } l = 1, ..., m$$

where $L_n = \frac{n^{4/5}}{k}$. We divide by L_n since the number of observations used to estimate a parameter increases at the rate $n^{4/5}$, instead of n. The expectation of $g_l(\gamma)$ is denoted by $g_{0,l}(\gamma)$ which decreases in γ_j for $j = 1, ..., M$

$$g_{0,l}(\gamma) = \frac{\sum_i}{L_n f_i} \{Ed_{i2}Ez_{i1} - Ed_{i1}Ez_{i2}\} \text{ for } l = 1, ..., M.$$

Its derivative

$$\frac{\partial g_{0,l}(\gamma)}{\partial \gamma_j} = \frac{\sum_i}{L_n} \{Ed_{i2}\frac{\partial Ez_{i1}}{\partial \gamma_j} - Ed_{i1}\frac{\partial Ez_{i2}}{\partial \gamma_j}\} < 0 \text{ for } j = 1, ..l - 1.$$

Note that $\frac{\partial Ez_{i2}}{\partial \gamma_j} = \frac{\partial Ez_{i1}}{\partial \gamma_j}$; further note that $Ed_{i1} < Ed_{i2}$ since d_{i1} is not artificially censored.

$$\begin{aligned} \frac{\partial g_{0,l}(\gamma)}{\partial \gamma_j} &= \frac{\sum_i}{L_n} \{Ed_{i2}\frac{\partial Ez_{i1}}{\partial \gamma_j} - Ed_{i1}\frac{\partial Ez_{i2}}{\partial \gamma_j}\} \\ &= \frac{\sum_i}{L_n} r\{-Ed_{i1}\frac{\partial Ez_{i2}}{\partial \gamma_j}\} < 0 \text{ for } j = l, ...M, \end{aligned}$$

since $\frac{\partial Ez_{i1}}{\partial \gamma_j}$ is not a function of γ_j for $j \geq l$. Therefore the following function is concave in γ_l for $l = l, ...M$:

$$Q_0 = -\sum_{l=1}^{M} \{g_{0,l}(\gamma)\}^2.$$

Theorem 4

Assume that we observe T spells for N individuals, that all the spells are independent and have the following hazards

$$\mu(t_{is}) = f_i \lambda_0(t).$$

The censoring time, c_{is}, is exogenous and we observe $y_{is} = \min(t_{is}, c_{is})$. Assume that $\lambda_0(t)$ is a bounded and continuous function of t. Artificial censoring and grouping is done as described above. Define $g_l(\gamma) = \frac{\sum_i}{L_n}\{\frac{d_{i1}z_{i2} - d_{i2}z_{i1}}{f_i}\}$ for $l = 1, ..., M$ and $Q_n = \sum_{l=1}^{M}\{g_l(\gamma)\}^2$. Maximizing $Q_n(\gamma)$ w.r.t. γ gives an estimate of $\hat{\lambda}_n(t)$ and this estimate converges in mean square to $\lambda_0(t)$, i.e.

$$\lim_{n \to \infty} E\{(\hat{\lambda}_n(t) - \lambda_0(t))^2\} = 0.$$

Proof: See appendix 5.

There is a similar theorem for the case in which we want to allow for regressors.

Theorem 5
Assume that we observe T spells for N individuals, that all the spells are independent given the regressors and have the following hazards

$$\theta(t_{is}) = f_i \lambda_0(t) e^{x_{is}\beta}.$$

The censoring time, c_{is}, is exogenous and we observe $y_{is} = \min(t_{is}, c_{is})$. The matrix ΔX has x_{is} as its columns. Assume that $\beta \in \Theta$ which is compact and $\Delta X' \Delta X$ has full rank. Further assume that $\lambda_0(t)$ is a bounded and continuous function of t. Artificial censoring and grouping is done as described above. Define $g_l(\gamma) = \frac{-1}{L_n} \sum \{\frac{d_{i2} z_{i1} - d_{i1} z_{i2}}{f_i}\}$ for $l = 1, \ldots, M$ where M is the number of intervals that increases with N. Define $g_r(\gamma) = \frac{-1}{L_n} \sum \{\frac{d_{i2} z_{i1} - d_{i1} z_{i2}}{f_i}\}$ for $l = 1, \ldots, K$ where K is the number of regressors that stays constant as $N \to \infty$. Finally, define $Q_n = -\{\sum_{l=1}^{M} \{g_l(\gamma)\}^2 + \sum_{r=1}^{K} \{g_r(\gamma)\}^2\}$.
Maximizing $Q_n(\gamma)$ w.r.t. β and γ gives an estimate of $\hat{\lambda}_n(t)$ and this estimate converges in mean square to $\lambda_0(t)$, i.e.

$$\lim_{n \to \infty} E\{(\hat{\lambda}_n(t) - \lambda_0(t))^2\} = 0,$$

while

$$\lim_{n \to \infty} E\{(\hat{\beta} - \beta_0)^2\} = 0.$$

Proof: See appendix 6.

In applications the number of intervals depends on the sample size. For a fixed sample size there is a trade-off between bias and variance of our estimation of the baseline hazard function. This is the usual trade-off that is very well discussed in the nonparametric literature: Using cross validation we choose the length of the intervals[9]. Cross validation techniques are discussed and reviewed in Härdle (1990).

Depending on the application, the relative importance of the baseline hazard compared to the regression coefficients differs. The two extreme interpretations are that either the baseline hazard or the regression coefficients[10] are nuisance parameters. In the case that the baseline hazard is interpreted as a nuisance parameter then the usual arguments for undersmoothing apply; see Powell (1994) for a general discussion.

3 Endogenous Censoring

In the last section we assume that all the spells are independent. This implies that the indicators for censoring are independent on the integrated hazards of other spells. The

[9] Or use cross validation to choose a number of observations per interval; let this number increase at a rate slower than N.

[10] The reason for introducing the regressors in this case is to control for covariates that differ across spells of the same individual.

assumption of independence is crucial for the theorems in that paragraph. To illustrate this, consider the case where $T = 2$ and z_{is} denotes the integrated hazard. We want the following expression to have zero expectation:

$$g_i = \frac{1}{\phi_i}\{d_{i2}z_{i1} - d_{i1}z_{i2}\}.$$

A sufficient condition is independence of the spell:

$$\begin{aligned} Eg_i &= \frac{1}{\phi_i}E\{d_{i2}z_{i1} - d_{i1}z_{i2}\} \\ &= \frac{1}{\phi_i}\{Ed_{i2}Ez_{i1} + Cov(d_{i2},z_{i1}) - Ed_{i1}Ez_{i2} - Cov(d_{i1},z_{i2})\}. \end{aligned}$$

We consider $Cov(d_{i2},z_{i1}) - Cov(d_{i1},z_{i2}) \neq 0$ to be a non-interesting case and therefore need $Cov(d_{i2},z_{i1}) = Cov(d_{i1},z_{i2}) = 0$ to apply the results of the last section.

However, the assumption of exogeneity is sometimes hard to justify. Consider two examples where the censoring time of the s^{th} spell of individual i, c_{is}, depends on the length of previous spells of individual i:

Example 1. Suppose that t_{is} denotes the time between two purchases of a product; the spells are independent of each other and are distributed exponentially with mean $\frac{1}{\phi_i e^{x_{is}\beta}}$, i.e.

$$t_{is} \sim \varepsilon(\phi_i e^{x_{is}\beta}).$$

Suppose we follow the customers for a six month period after their first purchase. So the censoring time for the first spell is six months, i.e. $c_{i1} = 6$. The censoring time for this second spell depends on the length of the first spell and is six months minus the duration of the first spell, i.e. $c_{i1} = 6 - t_{i1}$. Note that when $t_{i1} \geq 6$ then $c_{i2} \leq 0$ and we cannot even observe a part of the second spell. This section develops an estimator that is consistent despite this endogeneity problem. But first we will look at an easier example where the data are such that $c_{i2} > 0$ for all individuals.

Example 2. Suppose that t_{is} denotes the waiting time between two events and that the stochastic process of t_{is} is such that $t_{is} \leq t_{max}$ where t_{max} is a constant. For example, t_{is} could be the waiting time between inspection of the fire alarm of a building and t_{max} is the maximum time between two inspections as specified by some regulation. Assume that we observe individual i for a period c^{Data}, where $c^{Data} > t_{max}$ then we always observe at least a part of the second spell. More precisely, the second spell is censored at $c_{i2}(t_{i1}) = c^{Data} - t_{i1}$. We can artificially censor the second spell at $c_2 = c^{Data} - t_{max}$. This artificial censoring removes the endogeneity of the censoring time of the second period: c_2 no longer depends on t_{i1}. Thus we are back in the framework of the last section with exogenous censoring times and therefore consistent inference is possible for a wide range of models.

In this section we show that we can use the moment functions of the last section in many cases. A realistic sample scheme would be a scheme in which we are able to follow an individual for a certain period of time. Let c^{Data} denote the length of this period. Note that the first spell can be longer than c^{Data}; i.e. $t_{i1} \geq c^{Data}$. In that case we do not observe a second spell for that individual. Let us, for now, concentrate on the first and possibly second spell for each individual. Suppose we artificially censor the first period at c_1 and, if available, the second period at c_2 where $c_2 \leq c^{Data}$. Let us use the following moment function:

$$g = \frac{1}{N} \sum_i g_i$$

where

$$g_i = d_{i2} s_{i1} - d_{i1} s_{i2} \tag{9}$$

and s_{is} denotes the integrated hazard divided by the fixed effect, i.e. $s_{is} = \frac{z_{is}}{f_i}$. For the first spell we observe whether or not the observation is censored at c_1. So we know the values of d_{i1} and y_{i1}^{Data} [11]. We may, however, not observe a second spell for each individual: Suppose that the first spell lasts longer than the time we can follow an individual, c^{Data}. In that case we do observe y_{i1}^{Data} and d_{i1} but do not observe anything about a possible second spell, i.e. y_{i2} and d_{i2} are unobserved. As equation 9 shows, it is actually not a problem that we do not observe y_{i2} since d_{i1} is zero in those cases. However, we still need a value for d_{i2}. Therefore, we have to replace it with an unbiased estimate that is not correlated with y_{i2}. That is, the estimator $\hat{d}_{i2}$ should satisfy the following conditions:

$$E\hat{d}_{i2} = Ed_{i2} \text{ and } Cov(\hat{d}_{i2}; s_{i2}) = 0.$$

To distinguish the artificial censoring times from the censoring times of the data, we denote the artificial censoring times by c_1 and c_2 and the censoring of the data by c^{Data} and c_{i2}^{Data}. So $c_{i2}^{Data} = c^{Data} - y_{i1}$. Using elementary probability rules, we can write the Ed_{i2} as follows

$$Ed_{i2} = E(d_{i2}|t_{i2} \leq c_2^{Data})P(t_{i2} \leq c_2^{Data}) + E(d_{i2}|t_{i2} > c_2^{Data})P(t_{i2} > c_2^{Data}).$$

Whether c_2 is larger or smaller than c_2^{Data} depends on the length of the first spell. If $c_2 \leq c_2^{Data}$ then

$$E(d_{i2}|t_{i2} \leq c_2^{Data})P(t_{i2} \leq c_2^{Data}) = Ed_{i2}$$

since $E(d_{i2}|t_{i2} > c_2^{Data}) = 0$. If $c_2 > c_2^{Data}$ then

$$E(d_{i2}|t_{i2} \leq c_2^{Data})P(t_{i2} \leq c_2^{Data}) = Ed_{i2}^{Data}$$

and

$$P(t_{i2} > c_2^{Data}) = 1 - Ed_{i2}^{Data}.$$

So an estimator for $E(d_{i2}|t_{i2} > c_2^{Data})$ would be helpful for deriving an estimator for $\hat{d}_{i2}$. Note that $\hat{d}_{i2}$ cannot be based on the lengths of other spells since we did not make assumptions

[11] Where the superscript denotes that the censoring is a property of the dataset; artificial censored spells are denoted by y_{is}.

about the heterogeneity. Fortunately, we can use $y_{i1}^{Data} \geq c_{i1}$. For the observations in which we did not observe d_{i2} we always have $y_{i1}^{Data} \geq c_{i1}$. The quantity $y_{i1}^{Data} - c_{i1}$ can give an indication about the size of the fixed effect: If $y_{i1}^{Data} - c_{i1}$ is large, then the fixed effect is likely to be small. We use $y_{i1}^{Data} - c_{i1}$ to predict d_{i2}; i.e. to predict the value of d_{i2} in case we could have observed t_{i2}. We define $\Delta s_1 = s(y_{i1}^{Data};\gamma) - s(c_{i1};\gamma)$ and $\Delta s_2 = s(c_{i2};\gamma) - s(y_{i2};\gamma)$. In the case that we observe $d_{i2}^{Data} = 1$ then there is no need for adjustment; we denote the observable indicator by d_{i2}^{Data}. In the case that $d_{i2}^{Data} = 0$ we make the following adjustment: If $\Delta s_1 < \Delta s_2$ then we use $\tilde{d}_{i2} = 1$ as an estimate for d_{i2}. In other words, we define $\tilde{d}_{i2}$ to be the following:

$$\tilde{d}_{i2} = d_{i2}^{Data} + (1 - d_{i2}^{Data})\,\mathrm{ind}$$

where ind is an indicator that equals one if $\Delta s_1 < \Delta s_2$ and zero otherwise. After constructing an estimator for d_{i2} we can base inference on the same moment function as we did in section 2.

Lemma 3

Suppose we can follow each individual for a period with length c_D. Define $g(\gamma) = \frac{1}{N}\sum_i (\tilde{d}_{i2}s_{i1} - d_{i1}s_{i2})$ where $s_{is} = \frac{1}{f_i}\int_0^{y_{is}} \mu(s;x)ds$ and $y_{is} = \min(y_{is}^{Data};c_{is})$. Assume that the spells are either independent across both individuals and spells or that the probability of the second spell is censored, $E(1 - d_{i2})$, does not depend on t_{i1}. Then

$$Eg(\gamma_0) = 0.$$

Proof: Similar to lemma 2 where $E\tilde{d}_{i2} = Ed_{i2}$ (see appendix 7).
Lemma 3 only provides one moment function. As we showed in section 2, we can increase the number of moments by multiplying by an exogenous vector or additional censoring. If the conditions of theorem 7 are satisfied then the moment function of lemma 3 provides consistent estimates.

Theorem 6

Define $Q_n(\gamma) = -g(\gamma)^2$. If $g(\gamma)$ is continuous in γ and either assumption I or II is satisfied,

I. $g(\gamma)$ is monotonic and $E\sup_\gamma ||g(\gamma)|| < \infty$;

II. $\gamma \in \Gamma$, which is compact; $Eg(\gamma) = 0$ only if $\gamma = \gamma_0$ and $E\sup_{\gamma\in\Gamma} ||g(\gamma)|| < \infty$;

then $\hat{\gamma} = \arg\max_\gamma Q_n(\gamma)$ is a consistent estimate for γ_0, i.e.

$$\hat{\gamma} = \arg\max_{\gamma} Q_n(\gamma) \to_p \gamma_0.$$

Proof:
$Eg(\gamma_0) = 0$ by lemma 3. If assumption I is satisfied then $g(\gamma)$ is monotonic and $Q(\gamma) = -g(\gamma)^2$ is concave in γ and therefore $EQ(\gamma)$ is uniquely maximized at the truth. Therefore, maximizing $Q_n(\gamma)$ gives a consistent estimate; see Newey and McFadden, 1994, Theorem 2.7. If assumption II is satisfied then the conditions of Newey and McFadden, 1994, Theorem 2.6 are satisfied and maximizing $Q_n(\gamma)$ gives a consistent estimate of γ.Q.E.D.

3.1 Exponential hazard model

Suppose we can follow each individual for a period with length c_D. We can multiply the moment function of lemma 3 by a vector function of exogenous variables to derive more moment conditions. As in section 2, we multiply by Δx_i, the difference in the regressors between the first and second spell. That is, we use the following moment vector function

$$g(\beta) = \frac{1}{N}\sum_i g_i(\beta)$$

where

$$g_i(\beta) = \Delta x_i (d_{i2}s_{i1} - d_{i1}s_{i2}).$$

The resulting estimator is consistent. In the next subsection we illustrate this estimator with a simulation and compare it to other moment estimators.

3.2 Simulation

The following simulation illustrates the importance of taking endogenous censoring into account. We estimate an exponential hazard model with three estimators. The first is an estimator based on the integrated hazard as derived in the last subsection. We denote it by $\hat{\beta}^I$. The second and third estimator are standard estimators that are based on the difference in logarithms of durations. That estimator is consistent in the case no censoring takes place ($c^{Data} = \infty$ in the table). In the case that a part of the data is censored we can either

- ignore the censoring, i.e. use y_{is} instead of t_{is}; we refer to this approach as "Diff. 1" and denote the resulting estimator by $\hat{\beta}^{D1}$;
- ignore the censored observations i.e. only use those observations for which $d_{is} = 1$; we refer to this approach as "Diff. 2" and denote the resulting estimator by $\hat{\beta}^{D2}$.

The simulated sample consists of $N = 5000$ individuals. There is only one explaining variable and the difference of that regressor, Δx_i, follows a standard normal distribution. The fixed effects are distributed uniformly between 0.8 and 1.2. We repeat the estimation a 1000 times. In the table below we report the average and the variance of the estimates.

c^{Data}	Cens., 1st spell	Cens., 2nd spell	Int, $\hat{\beta}^I$	$\mathrm{var}\,\hat{\beta}^I$	Diff.1, $\hat{\beta}^{D1}$	$\mathrm{var}\,\hat{\beta}^{D1}$	Diff.1, $\hat{\beta}^{D2}$	$\mathrm{var}\,\hat{\beta}^{D2}$
∞	0%	0%	0.99995	$3.92 \cdot 10^{-4}$	0.99992	$6.53 \cdot 10^{-4}$	0.99992	$6.53 \cdot 10^{-4}$
10	0.0074%	3.23%	0.99992	$5.11 \cdot 10^{-4}$	0.9733	$6.07 \cdot 10^{-4}$	0.9304	$7.10 \cdot 10^{-4}$
5	0.871%	12.6%	1.00075	$6.57 \cdot 10^{-4}$	0.91860	$6.06 \cdot 10^{-4}$	0.85824	$8.40 \cdot 10^{-4}$
1	38.9%	73.6%	1.00009	$18.4 \cdot 10^{-4}$	0.67314	$8.54 \cdot 10^{-4}$	0.52343	$23.8 \cdot 10^{-4}$
0.5	62.2%	89.9%	0.99860	$29.5 \cdot 10^{-4}$	0.5459	$19.2 \cdot 10^{-4}$	0.3447	$52.3 \cdot 10^{-4}$

As Van den Berg (1999) notes, standard methods are indeed "particularly sensitive to censoring".

3.3 Piecewise-constant hazard

Assume that the government provides unemployment benefits for u_1 periods and that we observe two spells for N individuals. Suppose that we want to test whether the hazard differs before and after the benefits end. As in section 2, we can test this by estimating a piecewise-constant hazard model. It has the following hazard:

$$\begin{aligned} \theta(t_{is}) &= f_i \quad \text{if } t_{is} \leq u_1 \\ &= f_i \gamma \quad \text{if } t_{is} > u_1. \end{aligned}$$

The integrated hazard equals

$$\begin{aligned} z_{is} &= \int_0^{t_{is}} f_i ds = f_i t_{is} \text{ if } t_{is} \leq u_1 \\ &= f_i u_1 + f_i \gamma (t_{is} - u_1) \text{ if } t_{is} > u_1. \end{aligned}$$

The econometrician can censor the second spell at u_1; the number of periods for which the government provides unemployment benefits

$$g(\gamma) = \frac{\sum_i}{N} \left\{ \frac{d_{i2} z_{i1} - d_{i1} z_{i2}}{f_i} \right\}.$$

The estimator based on this moment function gives a consistent estimate since $Eg(\gamma_0) = 0$ is uniquely solved for γ.

3.4 Weibull Model

Suppose we can follow N individuals for a period with length c_D and that we want to estimate a Weibull model. The hazard has the form $\theta_{is} = \alpha f_i t_{is}^{\alpha - 1}$ $(s = 1, 2)$ and the integrated hazard is $z_{is} = f_i t_{is}^{\alpha}$ $(s = 1, 2)$. We observe $y_{is} = \min(t_{is}, c_{is})$. Theorem 6 suggests the following moment function

$$\begin{aligned} g(\gamma) &= \frac{1}{N f_i} \sum_{i=1}^{N} \{d_{i2} z_{i1} - d_{i1} z_{i2}\} \\ &= \frac{1}{N} \sum_{i=1}^{N} \{d_{i2} y_{i1}^{\gamma} - d_{i1} y_{i2}^{\gamma}\}. \end{aligned}$$

The equation $Eg(\gamma) = 0$ has a unique solution if $g(\gamma)$ is monotonic. Monotonicity of $g(\gamma)$ can always be assured by choosing a c_{i2} that is smaller than c_{i1}. As in section 2.2.4, we can combine $g(\gamma)$ with the moment function of the exponential hazard model to estimate a Weibull model that allows for regressors.

3.5 Exponential hazard model, alternative approach

An alternative approach to the problem of endogenous censoring is to concentrate on the observations of which neither the first nor the second spell are censored. In section 2 we used the fact that for uncensored spells, the expectation of the integrated hazard function equals one. Now we can use the expectation of the integrated hazard multiplied by its censoring indicator to derive an estimator. We define the normalized censoring time, $z_{is\max}$, as the integrated hazard evaluated at the censoring time for that spell, i.e.

$$z_{is\max} = \int_0^{c_{is}} \theta(t, x)dt.$$

Lemma 4 The expectation of the integrated hazard, multiplied by the indicator that indicated not being censored, $E(z_{is}d_{is})$, equals $1 - z_{is\max}e^{-z_{is\max}}$ where $z_{is\max} = \int_0^{c_{is}} \theta(t, x_{is})dt$, i.e.

$$E(z_{is}d_{is}) = 1 - z_{is\max}e^{-z_{is\max}} - e^{-z_{is\max}}.$$

Proof: See appendix 8.

For the exponential hazard model the integrated hazard and the normalized censoring time have the following form:

$$\begin{aligned} z_{is} &= \int_0^{y_{is}} f_i e^{x_{is}\beta} ds = f_i e^{x_{is}\beta} y_{is} \\ z_{is\max} &= f_i e^{x_{is}\beta} c_{is}. \end{aligned}$$

We want the same normalized censoring time for both spells

$$z_{i1\max} = z_{i2\max}.$$

Thus

$$f_i e^{x_{i1}\beta} c_{i1} = f_i e^{x_{i2}\beta} c_{i2}$$

and therefore

$$e^{x_{i1}\beta} c_{i1} = e^{x_{i2}\beta} c_{i2}.$$

This gives the censoring time for the first period

$$c_{i1} = e^{(x_{i1}-x_{i2})\beta} c_{i2}.$$

We artificially censor the second spells according to $c_{i1} = e^{(x_{i1}-x_{i2})\beta} c_{i2}$ and define

$$g(\beta) = (x_{i2} - x_{i1})\{e^{x_{i1}\beta} y_{i1} d_{i1} d_{i2} - e^{x_{i2}\beta} y_{i2} d_{i2} d_{i2}\} \text{ and}$$

$$\begin{aligned} Eg(\beta_0) &= \frac{(x_{i2} - x_{i1})}{f_i}\{E(z_{i1}d_{i1}d_{i2}) - E(z_{i2}d_{i2}d_{i1})\} \\ &= \frac{(x_{i2} - x_{i1})}{f_i}\{\{E(z_{i1}d_{i1}|d_{i2} = 1)\}\Pr(d_{i2} = 1) - \{E(z_{i2}d_{i2}|d_{i1} = 1)\}\Pr(d_{i1} = 1)\} \\ &= \frac{(x_{i2} - x_{i1})}{f_i}\{\{E(z_{i1}d_{i1}|d_{i2} = 1)\}\Pr(d_{i2} = 1) - \{E(z_{i2}d_{i2}|d_{i1} = 1)\}\Pr(d_{i1} = 1)\}. \end{aligned}$$

Since the normalized duration times are equal to each other, $z_{i1\max} = z_{i2\max}$, the probabilities of not being censored are equal too, i.e. $\Pr(d_{i2} = 1) = \Pr(d_{i2} = 1) = 1 - e^{-z_{i2\max}}$. The conditional expectations, $E(z_{i1}d_{i1}|d_{i2} = 1)$ and $E(z_{i2}d_{i2}|d_{i1} = 1)$ equal the unconditional expectations, i.e. $E(z_{i1}d_{i1}|d_{i2} = 1) = E(z_{i1}d_{i1})$ and $E(z_{i2}d_{i2}|d_{i1} = 1) = E(z_{i2}d_{i2})$ and therefore

$$\begin{aligned} Eg(\beta_0) &= \frac{(x_{i2} - x_{i1})}{f_i}\{1 - z_{i1\max}e^{-z_{i1\max}} - e^{-z_{i1\max}} - 1 - z_{i2\max}e^{-z_{i2\max}} - e^{-z_{i2\max}}\} \\ &= 0 \text{ using Lemma 3 and } z_{i1\max} = z_{i2\max}. \end{aligned}$$

We can also evaluate $Eg(\beta)$ at values other than the truth:

$$\begin{aligned} Eg(\beta) &= (x_{i2} - x_{i1})\{e^{x_{i1}\beta}Ey_{i1}d_{i1} - e^{x_{i2}\beta}Ey_{i2}d_{i2}\} \\ &= \frac{(x_{i2} - x_{i1})}{f_i}\{e^{x_{i1}\beta}Ef_iy_{i1}d_{i1} - e^{x_{i2}\beta}Ef_iy_{i2}d_{i2}\} \\ &= \frac{(x_{i2} - x_{i1})}{f_i}\{e^{x_{i1}(\beta-\beta_0)}Ef_ie^{x_{i1}\beta_0}y_{i1}d_{i1} - e^{x_{i2}(\beta-\beta_0)}Ef_ie^{x_{i2}\beta_0}y_{i2}d_{i2}\} \\ &= \frac{(x_{i2} - x_{i1})}{f_i}\{e^{x_{i1}(\beta-\beta_0)}E(z_{i1}d_{i1}) - e^{x_{i2}(\beta-\beta_0)}E(z_{i2}d_{i2})\}. \end{aligned}$$

Above we showed that $E(z_{is}d_{is}) = 1 - z_{is\max}e^{-z_{is\max}}$. We censored the second spell such that $z_{i1\max} = z_{i2\max}$ and, therefore, $E(z_{i1}d_{i1}) = E(z_{i2}d_{i2})$. Thus

$$\begin{aligned} Eg(\beta) &= \frac{(x_{i2} - x_{i1})}{f_i}\{e^{x_{i1}(\beta-\beta_0)}E(z_{i1}d_{i1}) - e^{x_{i2}(\beta-\beta_0)}E(z_{i1}d_{i1})\} \\ &= \frac{(x_{i2} - x_{i1})}{f_i}\{e^{x_{i1}(\beta-\beta_0)} - e^{x_{i2}(\beta-\beta_0)}\}E(z_{i1}d_{i1}). \end{aligned}$$

The derivative of $Eg(\beta)$ w.r.t. β is also of interest:

$$\begin{aligned} \frac{\partial Eg(\beta)}{\partial \beta} &= Eg_{\beta}(\beta) \text{ since the boundaries of } t_{is} \text{ do not depend on } \beta \\ &= \frac{(x_{i2} - x_{i1})}{f_i}\{x_{i1}e^{x_{i1}(\beta-\beta_0)} - x_{i2}e^{x_{i2}(\beta-\beta_0)}\}E(z_{i1}d_{i1}). \end{aligned}$$

Since we have normalized x_{is}, $\sum x_{is} = 0$, $x_{i2} = -x_{i1}$ and therefore

$$\begin{aligned} \frac{\partial Eg(\beta)}{\partial \beta} &= \frac{-2x_{i1}}{f_i}\{x_{i1}e^{x_{i1}(\beta-\beta_0)} + x_{i1}e^{-x_{i1}(\beta-\beta_0)}\}E(z_{i1}d_{i1}) \\ &= -\frac{2x_{i1}^2}{f_i}\{e^{x_{i1}(\beta-\beta_0)} + e^{-x_{i1}(\beta-\beta_0)}\}E(z_{i1}d_{i1}) < 0 \end{aligned}$$

since $x_{i1}^2 > 0$, $f_i > 0$ and $e^{x_{i1}(\beta-\beta_0)} + e^{-x_{i1}(\beta-\beta_0)} > 0$. The column vector $Eg(\beta)$ has as many rows as x_{is}. When we differentiate the j^{th} element of column $Eg(\beta)$ then the result above helps to show that β_0 is a locally unique solution to $Eg(\beta) = 0$.

Theorem 7
Assume that the hazard has the following form

$$\lambda_{is} = f_i e^{x_{is}\beta}.$$

We define the function $k(t_{i1})$ as a function of the length of the first spell that determines the censoring time c_{i2}. The censoring time of the second spell may depend on the length of the first spell, i.e. $c_{i2} = k(t_{i1})$. We will ignore spells that we censored in the first period. We define the function $k(t_{i1})$ as a function of the length of the first spell that determines the censoring time c_{i2}: $c_{i2} = k(t_{i1})$. Assume $\beta \in \Theta$; which is compact.
We censor the first spells according to $c_{i1} = e^{-(x_{is} - x_{is})\beta} c_{i2}$ and define

$$g(\beta) = (x_{i2} - x_{i1})\{e^{x_{i1}\beta} y_{i1} d_{i1} d_{i2} - e^{x_{i2}\beta} y_{i2} d_{i2} d_{i1}\}$$

$$Q(\beta) = -g(\beta)' g(\beta).$$

Maximizing $Q(\beta)$ w.r.t. β gives a consistent estimate for β; i.e.

$$\hat{\beta} = \arg\max_{\beta} Q(\beta) \to_p \beta_0 \text{ for } N \to \infty.$$

Proof: See appendix 9.

4 Endogenous regressors: Tools to estimate hysteresis

In the previous sections, we assumed that the regressors, x_{is}, were exogenous. This section provides tools to estimate duration models in which the regressors, x_{is}, can depend on the length of previous spells. Following Honoré (1993) we refer to such regressors as endogenous regressors. Honoré (1993) notes that "the main problem with identification of multispell models is that if the model has 'lagged duration dependence', then one of the regressors (lagged duration) will be endogenous". He gives results for exogenous regressors and further notes sensitivity to the mixing distribution. There are several reasons why one may want to allow for the endogeneity of the regressors. The first is that the individual may choose x_{is} given the fixed effect f_i and the length of the previous spells. Mundlak (1961) argues that the value of the fixed effect f_i influences the choice of regressor x_{is} and that therefore f_i and x_{is} are dependent. Similar reasoning gives that x_{is} may be chosen by the individual as a function of both f_i and the length of previous spells. Another reason for endogeneity might be that the government is more likely to offer job training programs to people with longer unemployment spells. In that case participation in a program can be endogenous for two reasons: selection of an agency and self selection of the individuals. Another regressor that is likely to depend on previous unemployment spells is the employment spell. Finally, the hazard in one spell may be determined by the length of previous spells in which case a function of the length of previous spells appears as a regressor. Whatever the mechanism that causes endogeneity, it seems to be interesting to have an estimator that is robust against endogeneity.

We first present how to handle endogenous regressors in the exponential hazard model. Then we discuss the estimation of hysteresis and present a more general framework that allows for endogenous regressors, duration dependence and endogenous censoring.

4.1 Exponential hazard model with endogenous regressors

Suppose that we observe two uncensored spells for N individuals and observe, for each spell, a vector of regressors w_{is}. Assume that the hazard has an exponential form and that we want to allow for the endogeneity of the regressors[12]. So the hazard has the following form:

$$\begin{aligned} \theta_{i1} &= f_i \\ \theta_{i2} &= f_i e^{\Delta w_i \beta} \end{aligned} \tag{10}$$

where $\Delta w_i = w_{i2} - w_{i1}$. The integrated hazards are $z_{i1} = f_i t_{i1}$ and $z_{i2} = f_i e^{\Delta w_i \beta} t_{i2}$ respectively. We can use the fact that the expectation of the difference between the first and second hazard equals zero and define

$$g_i(\beta) = t_{i1} - e^{\Delta w_i \beta} t_{i2}.$$

Example 2 of section 2.1 deals with exogenous regressors and multiplies $g_i()$ by the difference in the regressors. However, if the elements of the vector Δw_i are correlated with t_{i1} then $E\,\Delta w_i g_i(\beta_0) \neq 0$ and therefore the approach of example 2 does not work. Fortunately, we can reparameterize the hazard rate such that the regressors are no longer correlated with t_{i1}. Let Δw_i^m denote the m^{th}element of the vector Δw_i, $m = 1, \ldots, M$ where M is the number of regressors. We define

$$\begin{aligned} \Delta w^m &= \{\Delta w_1^m, \Delta w_2^m, \ldots, \Delta w_{(N-1)}^m, \Delta w_N^m\} \text{ and} \\ t_1 &= \{t_{11}, t_{21}, \ldots, t_{(N-1)1}, t_{N1}\}. \end{aligned}$$

We regress Δw^m on t_1 using ordinary least squares and obtain an estimate for the slope parameter, α_s^m; and an estimate for the constant, α_c^m. A transformation of Δw_i^m that is not correlated with t_{i1} is, obviously

$$\tilde{w}_i^m = \Delta w_i^m - \alpha_s^m t_{i1}.$$

An equivalent definition of $\tilde{w}^m$ uses the OLS residuals[13] e_i^m :

$$\tilde{w}_i^m = \alpha_c^m + e_i^m.$$

The new variable, $\tilde{w}_i^m$, is not correlated with t_i since $\sum_i e_i^m t_{i1} = 0$. We can multiply $g_i(\beta)$ by $\tilde{w}_i^m$ and obtain

$$g^*(\beta) = \frac{1}{N} \sum_i \tilde{w}_i^m g_i(\beta).$$

Since $\tilde{w}_i^m$ is not correlated with $g_i(\beta_0)$ we have the following result, (see appendix 11 for details),

$$E g^*(\beta_0) = \frac{1}{N} E \sum_i \tilde{w}_i^m g_i(\beta) = 0.$$

---

[12] Effects of the possible spells prior to the first spell are incorporated in the fixed effect.

[13] Using the identity $\Delta w_i^m = \alpha_c^m + \alpha_s^m t_{i1} + e_i^m$.

Identification is assured by a full rank condition and the use of the moment function $g(\beta) = \frac{1}{N}\sum_i g_i(\beta)$ in addition to the moments implied by $g^*(\beta)$. These two moment functions identify a model that is less restrictive than (10). This less restrictive model has an exponential duration dependence in the sense that the hazard of the second period is an exponential function of the first spell:

$$\begin{aligned} \theta_{i1} &= f_i \\ \theta_{i2} &= f_i e^{\tilde{w}_i \beta + \delta t_{i1}}. \end{aligned} \tag{11}$$

Note that $\tilde{w}_i$ is not correlated t_{i1} and that we can estimate this hazard rate by stacking the moment functions of theorem 1 and 2. Details are given in the proof of theorem 8. Define $\alpha = \{\alpha_s^1, \alpha_s^2, ..., \alpha_s^M\}$; if $\delta = \beta'\alpha$ then

$$\tilde{w}_i \beta + \delta t_{i1} = \tilde{w}_i \beta + \beta' \alpha t_{i1} = \Delta w_i \beta.$$

Therefore, if δ is restricted to equal $\beta'\alpha$ then model (10) and (11) are equivalent. Therefore, identification of model (11) implies that model (10) is identified as well. The next theorem uses this argument to prove identification and shows how to estimate the exponential hazard model with endogenous regressors.

Theorem 8

Suppose that we observe two uncensored spells for N individuals and a vector of regressors w_{is} for each spell. Assume that the hazard regressors can depend on the length of previous spells and that the hazard has the following form:

$$\begin{aligned} \theta_{i1} &= f_i \\ \theta_{i2} &= f_i e^{\Delta w_i \beta} \end{aligned}$$

where $\Delta w_i = w_{i2} - w_{i1}$. Further assume that the matrix with rows $\{t_{i1}; \Delta w_i'\}$ has full column rank and that $\beta \in \Theta$, which is compact. Let $\tilde{w}_i^m$, $g^*(\beta)$ and $g(\beta)$ be defined as above. Further define

$$Q(\beta) = -\{g(\beta)^2 + g^*(\beta)' g^*(\beta)\}.$$

Maximizing $Q(\beta)$ w.r.t. β gives a consistent estimate for β, i.e.

$$\hat{\beta} = \arg\max_{\beta} Q(\beta) \to_p \beta_0 \text{ for } N \to \infty.$$

Proof: See appendix 12.

In the case that the number of spells is larger than two for all individuals we can pair them, i.e. equate the integrated hazard of the first spell to the integrated hazard of the second spell and, in a second equation, equate the integrated hazard of the second spell to the third, etc. An alternative approach is to avoid making pairs and regress the endogenous variables on t_{i1}, t_{i2}.

Before we look at more general theorems with endogenous regressors we discuss the concept of unemployment hysteresis. The hysteresis literature in macro economics implies that the hazard rate for the unemployed is path dependent. We think that hysteresis is an important subject in itself. Moreover, its estimation can illustrate the generalizations of theorem 8.

4.2 Hysteresis

Originating in physics, the word hysteresis is used to describe the circumstance in which the equilibrium of a system depends on the history of that system. In economics it is most commonly used to consider the natural rate of unemployment, where the equilibrium is said to be path-dependent, i.e. it depends on the actual history or path of unemployment. Hysteresis may result in a rise of the proportion of long term unemployment to the total unemployment, where the long term unemployed have a very low probability of ¯nding a job. This low hazard rate for the long term unemployed is caused by a lack of relevant skills and psychological reasons[14]. Blanchard and Wolfers (1999) argue that hysteresis can at least partly explain the high unemployment rate in Europe but the existence of the phenomenon of hysteresis is not commonly accepted. Since the policy implications are large it is nice to perform an econometric test. The data for the test may have no transition for some individuals and may be right censored as well. A good test for path dependence of employment and unemployment should take care of the employment spells. We assume that the employment in°uences the hazard of the next unemployment spell through the function b(s) where s is the length of the employment spell. Heckman and Borjas (1980) discuss duration dependence and lagged duration dependence in the framework of a random e®ects model. Heckman (1991) notes that such models are very sensitive to the assumed distribution of the random e®ect. So a °exible form of the heterogeneity as well as endogenous regressors are desirable properties of an estimator of hysteresis. Before we extent the estimator of the last section we discuss an easy example of lagged duration dependence with just one endogenous regressor.

Example: Lagged duration dependence in an one-parameter model.
Consider the following model of lagged duration dependence where we observe two spells, possible censored, for each individual. The spells are independent across individuals and have the following hazards:

$$\mu_{i1} = f_i$$
$$\mu_{i2} = f_i e^{\gamma r(y_{i1}^{Data})}$$

where $r(y_{i1}^{Data}) > 0$ and either $r()$ ¯nite or $r(y_{i1}^{Data})$ bounded in probability. Note that, in the presence of lagged duration dependence, the probability of the second spell being censored depends on the length of the ¯rst spell, even if the censoring times are exogenous. The last section suggested to remove the endogeneity by arti¯cial censoring. So we use $y_{i1} = \min(y_{i1}^{Data}; c_{i1})$ and $y_{i2} = \min(y_{i2}^{Data}; e^{-\gamma r(y_i)} c_{i2})$. The integrated hazards have the

[14] The lack of relevant skills and the psychological problems do not indicate hysteresis per s¶ since both might be caused by a person speci¯c e®ect.

following form:

$$
\begin{aligned}
z_{i1} &= \int_0^{y_{i1}} \theta_{i1} ds = f_i y_{i1} \\
z_{i2} &= \int_0^{y_{i2}} \theta_{i2} ds = f_i e^{\gamma r(y_i)} y_{i2}.
\end{aligned}
$$

Note that, after artificial censoring, the probability of the second spell being censored does not depend on the length of the first spell anymore:

$$
\begin{aligned}
Prob(2^{nd}\text{ spell censored}) &= E(1 - d_{i2}) = e^{-z_{max}} \text{ where} \\
z_{max} &= f_i e^{\gamma r(y_i)}(e^{-\gamma r(y_i)} c_{i2}) = f_i c_{i2}.
\end{aligned}
$$

After the endogeneity is removed we can use the following moment function:

$$
g(\gamma) = \frac{1}{N f_i} \sum_i \{d_{i2} z_{i1} - d_{i1} z_{i2}\} = \frac{1}{N f_i} \sum_i \{d_{i2} y_{i1} - d_{i1} e^{\gamma r(y_i)} y_{i2}\}.
$$

According to theorem 6 the resulting estimator is consistent (see Appendix 12 B for details).

The models are usually more complicated than this example. In particular, models about hysteresis in unemployment data should allow for several endogenous regressors and a duration dependent hazard. The duration dependence should be more flexible than the Weibull model since monotonicity of the baseline hazard is a rather strong assumption[15]. We are therefore primarily interested in the piecewise-constant hazard model.

4.3 Piecewise-constant hazard with endogenous regressors

Suppose we observe two uncensored spells for N individuals and want to estimate a piecewise-constant hazard model with endogenous regressors. Assume that the hazard has the following form

$$
\begin{aligned}
\theta_{i1} &= f_i \pi(t) \\
\theta_{i2} &= f_i e^{\Delta w_i \beta} \pi(t)
\end{aligned}
\qquad (12)
$$

where $\pi(t)$ denotes the baseline hazard and $\Delta w_i = w_{i2} - w_{i1}$. Assume that the piecewise-constant hazard has three breaking points and that these breaking points are at u_1, u_2 and u_3. This means that $\pi(t)$ can have four different values over four different intervals:

$$
\begin{aligned}
\pi(t) &= 1 \text{ if } 0 \le t < u_1 \\
\pi(t) &= \gamma_1 \text{ if } u_1 \le t < u_2 \\
\pi(t) &= \gamma_2 \text{ if } u_2 \le t < u_3 \\
\pi(t) &= \gamma_3 \text{ if } u_3 \le t < \infty.
\end{aligned}
$$

[15] For example, the baseline hazard is likely to `peak' around the time that people are no longer eligible for unemployment benefits.

Since the model is more complicated then the previous one we estimate it in different steps. In the first step we either ignore the duration dependence or the endogeneity of some indicators and estimate a simplified model. In the latter steps we pick up what was ignored in the first step.

Ignore duration dependence

Suppose we choose to ignore the duration dependence at first. We can then proceed with the following steps:

1. Ignoring the duration dependence brings us back in the framework of theorem 8: We regress[16] the endogenous regressors Δw_i on t_1 and obtain $\hat{w}^m$:

$$\hat{w}_i^m = \alpha_c^m + e_i^m.$$

   The initial estimate for the regressors is denoted by β_1.

2. We focus on the interval $t \in [0, u)$ and therefore censor the first and second spell at c_{i1} and c_{i2} respectively where c_{i1}, $c_{i2} < u_1$. Following the reasoning of the last section we want to remove the dependence between the first and second period and, in particular, remove the dependence between the length of the first spell, t_{i1}, and the expectation of d_{i2}. We therefore choose $c_{i1} = 0.8u_1$ and $c_{i2} = e^{\Delta w_i \beta_1} k_1$ such that $c_{i2} < u_1 \; \forall i$. Note that this is the framework with an exponential hazard model and that the expectation of Ed_{i2} does not depend on t_{i1}. Define $y_{is} = \min(t_{is}, c_{is})$. We also remove the endogeneity of the endogenous regressors by regressing Δw_i on y_1 and d_{i1} to obtain $\hat{w}^m$. We can use the moment functions that are suggested by theorem 1 and 2:

$$\begin{aligned} g(\beta) &= \frac{1}{N} \sum_i d_{i2} y_{i1} - d_{i1} e^{\Delta w_i \beta} y_{i2} \\ g^*(\beta) &= \frac{1}{N} \sum_i \hat{w}^m (d_{i2} y_{i1} - d_{i1} e^{\Delta w_i \beta} y_{i2}). \end{aligned}$$

   Note that d_{is} and y_{is} are functions of β since $c_2 = e^{\Delta w_i \beta_1} k_1$. We denote the resulting estimate by $\hat{\beta}$ and note that it is a consistent estimate for β of equation (12).

3. After obtaining a consistent estimate for β we focus on the γ's: Theorem 4 and 5 suggest moments to estimate these parameters in the case of exogenous censoring[17]. As in step 2 we can remove the endogeneity by artificially censoring the data. For example, censor the first spell at u_2 and the second sample at $c_{i2}(\Delta w_i \beta_1, k_2)$ where k_2 is chosen s.t. c_{i2} is smaller than u_1 for 80% of the population. Note that in theorem 4 and 5 the artificial censoring is at u_1 and u_2; however, since we have to remove the endogeneity of the censoring we cannot do that anymore. Suppose the baseline hazard has L parameters, then we derive the moment functions for the other parameters in a

[16] If the econometrician is willing to assume that some of the covariates are exogenous then $\hat{w}^m = \Delta w_i$ for those covariates.

[17] Note, however, that the dimension of γ is fixed here.

similar way

$$g_l^{PC}(\beta;\gamma) = \frac{1}{N}\sum_i d_{i2}^l y_{i1}^l - d_{i1}^l e^{\Delta w_i \beta} y_{i2}^l \text{ for } l = 1,\ldots,L.$$

Note that d_{is} and y_{is} are functions of the parameters since $c_1^l = u_{l+1}$ and $c_{i2}^l(\Delta w_i \beta_1; k_2^l)$ where k_2^l is chosen s.t. 80% of the second spell we have $c_{i2}^l(\Delta w_i \beta_1; k_2^l) < u_l$.

4. The final step is to estimate all the parameters simultaneously. We increase k_1 s.t. $c_{i2} = e^{\Delta w_i \beta_1} k_1 < u_1$ for 80% of the spells and define

$$\begin{aligned} g^E(\beta;\gamma) &= \frac{1}{N}\sum_i d_{i2}s_{i1} - d_{i1}s_{i2} \\ g^{*E}(\beta;\gamma) &= \frac{1}{N}\sum_i \hat{w}^m (d_{i2}s_{i1} - d_{i1}s_{i2}) \end{aligned}$$

where s_{is} denotes the integrated hazard divided by its fixed effect, i.e. $s_{is} = \frac{z_{is}}{f_i}$ and $\hat{w}^m$ is obtained by regressing Δw_i on t_1 and d_{i1}. We further define $g^{PC}(\beta;\gamma) = \{g_1^{PC}(\beta;\gamma), g_2^{PC}(\beta;\gamma), \ldots, g_L^{PC}(\beta;\gamma)\}$. At the true value, $\{\beta_0;\gamma_0\}$ all the moments have zero expectation:

$$\begin{aligned} Eg^{PC}(\beta_0;\gamma_0) &= 0 \\ Eg^{E}(\beta_0;\gamma_0) &= 0 \\ Eg^{*E}(\beta_0;\gamma_0) &= 0. \end{aligned}$$

Theorem 9 uses this overidentification system of equation to estimate the parameters consistently. The uniqueness of the solution to this system of equations follows from theorems 1, 2, 5 and 8. From theorem 1, 2 and the discussion of theorem 8 we conclude that, if $c_{i2} < u_1 \ \forall i$ [18] then $Eg^E(\beta) = Eg^{*E}(\beta) = 0$ is uniquely solved for $\beta = \beta_0$. Using arguments of theorem 5 we conclude that the estimation of the finite parameter vector γ is possible. We finally conclude that consistency by allowing a low percentage of c_{i2} to be larger then u_i.

Instead of estimating a simplified model first one can also choose to do the estimation of step 4 directly. If one feels like having a preliminary step, one can also choose to ignore the endogeneity of spells.

Ignore endogeneity for short spells

1. Censor the first and second spell at u_1; regress the endogenous regressors Δw_i on y_1 and d_{i1} to obtain $\hat{w}^m$. Define $y_{is} = \min(t_{is}; u_1)$ and use the moment functions that are suggested by theorem 1 and 2:

$$\begin{aligned} g(\beta) &= \frac{1}{N}\sum_i d_{i2}y_{i1} - d_{i1}e^{\Delta w_i \beta} y_{i2} \\ g^*(\beta) &= \frac{1}{N}\sum_i \hat{w}^m (d_{i2}y_{i1} - d_{i1}e^{\Delta w_i \beta} y_{i2}). \end{aligned}$$

[18] So that $g^E(\beta;\gamma) = g^E(\beta)$ and $g^{*E}(\beta;\gamma) = g^{*E}(\beta)$.

The estimate that follows from this is denoted by β_1. Note that we did not correct for the correlation between d_{i2} and y_{i1} or the between d_{i1} and y_{i2}.

2. Use the initial estimate β_1 and follow the scheme from step 2 above.

We summarize the findings of this section in a theorem:

Theorem 9
Suppose we observe two uncensored spells for N individuals and that the hazard rate is piecewise-constant with endogenous regressors:

$$\begin{aligned}\theta_{i1} &= f_i\lambda(t)\\ \theta_{i2} &= f_i e^{\Delta w_i \beta}\lambda(t)\end{aligned}$$

where $\Delta w_i = w_{i2} - w_{i1}$. Assume that the piecewise-constant hazard has L breaking points and that these breaking points are at $u_1; u_2; \ldots u_L$. Assume that $\{\beta; \gamma\} \in \Theta$; which is compact. Let $g^{PC}(\beta;\gamma)$, $g^{E}(\beta;\gamma)$ and $g^{*E}(\beta;\gamma)$ be defined as above. The system of equations $Eg^{PC}(\beta;\gamma) = Eg^{E}(\beta;\gamma) = Eg^{*E}(\beta;\gamma) = 0$ is uniquely solved for $\{\beta_0; \gamma_0\}$ and the resulting GMM estimator is consistent. See appendix 13 for details.

For hazard functions with other baseline hazard function we can follow the similar technique. Let the parameters of be denoted by θ. For uncensored spells we have the following moment function

$$g(\theta) = \frac{1}{N}\sum_i (s_{i1} - s_{i2})$$

Depending on the specification of the baseline hazard we need to generate more moments by artificial censoring. The derivation of the moments of the regressors is a close analogue to theorem 9, the only difference being that, for the endogenous regressors, Δw_i is now regressed on s_1. Following the notation of theorem 9, we can define $\tilde{w}_i^m = \alpha_c^m + e_i^m$ and use the following moment functions:

$$g^E(\theta) = \frac{1}{N}\sum_i \tilde{w}_i(s_{i1} - s_{i2}).$$

In case we want to use multiply a function of the regressors with moments that are based on artificial censored data then we need to regress Δw_i on s_1 and d_1; where, with an abuse of notation, s_1 denotes the integrated hazard of the possibly censored spell. Note that the artificial censoring should be such that the probabilities of censoring should not be a function of the endogenous elements of Δw_i. In that case

$$\begin{aligned}E_{t_1;t_2}(d_{i2}s_{i1} - d_{i1}s_{i2}) &= E_{t_1}\{E_{t_2}(d_{i2}s_{i1} - d_{i1}s_{i2})|t_1\}\\ &= Ed_{i2}Es_{i1} - Ed_{i1}Es_{i2} = 0\end{aligned}$$

Following previous notation we define $\tilde{w}_i^m$ as a constant plus the residuals from regressing Δw_i on s_1 and d_1:

$$\tilde{w}_i^m = \alpha_c^m + e_i^m.$$

Note that e_i^m is not correlated with $d_{i2}s_{i1} - d_{i1}s_{i2}$ and therefore

$$E\frac{1}{N}\sum_i w_i^m(d_{i2}s_{i1} - d_{i1}s_{i2}) = 0.$$

So the integrated hazard provides moment functions for duration models with endogenous regressors and makes consistent inference possible for a wide range of models with uncensored observations.

4.3.1 Piecewise-constant hazard with endogenous censoring and endogenous regressors

In section 3 we derived estimators for duration models with endogenous censoring. The main insight was to remove the endogeneity of the censoring of the second period. Therefore, the line of thought and algebra of section 3 is very close to the last section where the endogeneity of the censoring is caused by regressors that depend on t_{i1}. Combining these two section gives a repetition of the last section were the arguments of theorem 1 and 2 are replaced by theorem 6 and the discussion of section4.1. The problem with endogenous censoring is that the second spell may not be observed at all. Therefore we need an estimator for the expectation of d_{i2}. A good estimator for Ed_{i2} is uncorrelated with y_{i1} and has an expectation equal to Ed_{i2} [19]. Following previous notation we denote this estimator by $\hat{d}_{i2}$. To assure the existence of a good estimator we may need to artificially censor the first period as is discussed in 3. After construction $\hat{d}_{i2}$ we can use exactly the same moment functions as in the last section.

Theorem 10
Suppose we follow N individuals for a period with length c^D. Assume that the hazard rate is piecewise-constant with exponential regressors:

$$\begin{aligned} \theta_{i1} &= f_i\pi(t) \\ \theta_{i2} &= f_i e^{\Delta w_i \beta}\pi(t) \end{aligned}$$

where $\Delta w_i = w_{i2} - w_{i1}$. Assume that the piecewise-constant hazard has L breaking points and that these breaking points are at $u_1; u_2; \ldots u_L$ and that $u_L < 0.5c^D$. Assume that $\{\beta; \gamma\} \in \Theta$, which is compact and that we can find an estimator $\hat{d}_{i2}$ is uncorrelated with y_{i1} and $E\hat{d}_{i2} = Ed_{i2}$. Let $g^{PC}(\beta;\gamma)$, $g^E(\beta;\gamma)$ and $g^{*E}(\beta;\gamma)$ be defined as above where $\hat{d}_{i2}$ replaces d_{i2}. The system of equations $Eg^{PC}(\beta;\gamma) = Eg^E(\beta;\gamma) = Eg^{*E}(\beta;\gamma) = 0$ is uniquely solved for $\{\beta_0; \gamma_0\}$ and the resulting GMM estimator is consistent. Proof: Similar to theorem 9.

Other hazard function can be dealt with by using the following moment function:

$$g(\theta) = \frac{1}{N}\sum_i (\hat{d}_{i2}s_{i1} - d_{i1}s_{i2}).$$

[19] where $y_{i1} = \min(y_{i1}^D; c_{i1})$ and d_{i2} is an indicator that equals one if the unobserved t_{i2} is larger than c_{i2}.

More moment function are generated by artificial censoring. Endogenous regressors are projected in the space spanned by s_1 and d_1 as was done in the last section; we define $\hat{w}_i^m = \hat{\alpha}_c^m + e_i^m$ and use the following moment vector function:

$$g^*(\theta) = \frac{1}{N}\sum_i \hat{w}_i^m (d_{i2}s_{i1} - d_{i1}s_{i2}).$$

As was shown in section 2, the moment functions that are suggested by the integrated hazard can deal with a wide range of models. Moreover, it can also deal with more realistic data than uncensored spells with exogenous regressors.

5 Single spell

In the previous sections we assume that we observe more than one spell for each individual. In this section we assume that we have only one spell for each individual. We therefore cannot allow for a fixed effect since the common parameters would be no longer identified. However, we can do something that is at least mathematically similar to the previous sections: Instead of having another spell of the same person we can use a weighted average of other individuals whose observed characteristics suggest that their values of the fixed effects may be similar. Powell (1987) develops this idea and Powell and Honoré (1998, PH from now on) apply it to a number of estimation procedures that are based on differencing out the fixed effect. PH make certain assumptions about how to compare the dependent variable of an individual with a weighted average of dependent variables of similar individuals. Under these, as well as under slightly weaker assumptions, the estimating functions of the earlier sections give consistent estimates of the common parameters. Note that these estimating functions were based on the integrated hazard and, therefore, do not belong to the class of "difference estimators". Indeed, an important motivation of this paper is the unsatisfactory assumptions of the difference estimator for the duration model[20]. Nevertheless, we find the framework of PH and Powell (1987) very useful and therefore follow their assumptions about matching of individuals.

Suppose that the fixed effect can be written as an unknown function of a vector of regressors w_i, $f_i = v(w_i)$. PH and Powell (1987) note that a feasible version of the idea of 'differencing away' uses all pairs of observations and gives bigger weight to pairs for which w_i is close to w_j. The weights are chosen in such a way that asymptotically, only pairs with $w_i - w_j$ in a shrinking neighborhood of 0, matter.

Assumption Kernel. K is bounded, differentiable with bounded derivative K', and of bounded variation. Furthermore, $\int K(u)du = 1$, $\int |K(u)|du < \infty$, and $\int |K(u)| \, \|\xi\| \, d\xi < \infty$.

The assumptions made on the kernel are satisfied for most kernels. As usual in semiparametric literature, the bandwidth shrinks with increasing n. When we compare the duration of

[20] If we take the difference in the logarithms of the durations then we need to observe these durations, i.e. no censoring of any kind. Another disadvantage is that the form of the baseline hazard is restricted to the Weibull.

individual i we compare it to individuals with w_j in the neighborhood of w_i; this neighborhood shrinks with increasing n. We can think of the neighborhood of w_i as an L dimensional ball. How fast the radius of this ball can shrink depends on the dimension of w_i as is shown in the following assumption about the bandwidth:

Assumption Bandwidth $h_n > 0$; $h_n = o(1)$ and $h_n^{-1} = O(n^{1/2L})$ where L is the dimension of w.

We make the following assumption on the data:

Assumption Data w_i is continuously distributed with bounded density.

These three assumptions enable us to modify Theorem 1 and Theorem 2. These theorems required panel data; here we give single spell variants. For theorem 1 we had the following estimating function:

$$g_l(\beta) = \frac{1}{N}\sum_{i=1}^{N}\Big\{\sum_s d_{is}\sum_s x_{isl}d_{is}e^{x_{is}\beta}y_{is} - \sum_s x_{is}d_{is}\sum_s d_{is}e^{x_{is}\beta}y_{is}\Big\} \text{ for } l = 1,\ldots,K.$$

In the case that $T = 2$ we could simplify this:

$$g_l(\beta) = \frac{1}{N}\sum_{i=1}^{N}(x_{i1} - x_{i2})\{e^{x_{i2}\beta'}y_{i2}d_{i1} - e^{x_{i1}\beta'}y_{i1}d_{i2}\} \text{ for } l = 1,\ldots,K.$$

Instead of using two spells of the same period we will now use a spell of some individual and the weighted average of a group:

$$g_l(\beta) = \binom{n}{2}^{-1}\sum_{j<i}\frac{1}{h_n^L}K\Big(\frac{w_i - w_j}{h_n}\Big)(x_i - x_j)\{e^{x_j\beta'}y_jd_i - e^{x_i\beta'}y_id_j\} \text{ for } l = 1,\ldots,K.$$

Theorem 1A.
Suppose Assumption Kernel, Assumption Bandwidth and Assumption Data hold. Assume $\beta \in \Theta$, which is compact. Define $Q(\beta) = -\sum_l g_l(\beta)^2$ where $g_l(\beta)$ is defined as above. Maximizing $Q(\beta)$ w.r.t. β gives a consistent estimate for β; i.e.

$$\hat\beta = \arg\max_{\beta} Q(\beta) \to_p \beta_0 \text{ for } N \to \infty.$$

Proof:
The function $g(\beta)$ is continuous in β. Let $Eg_{2l}(\beta)$ for $l = 1,\ldots,K$, denote the expectation of the estimating function of theorem 1 and $Q_{2,0} = -\sum_l g^2_{2l}(\beta)^2$ denote the objective function of theorem 1. Now $g(\beta) \to^U Eg_{2l}(\beta)$ and $Q(\beta) \to^U Q_{2,0}$. The objective function $Q_{2,0}$ is maximized at the truth (see appendix 1) and consistency of $\hat\beta$ follows.

In section 2 we derived a general theorem for single parameter hazards. When assumptions Kernel, Bandwidth and Data are mode we can derive a similar theorem for single spell data. Define:

$$g(\gamma) = \binom{n}{2}^{-1} \sum_{j<i} \frac{1}{h_n^L} K\left(\frac{w_i - w_j}{h_n}\right)\{s_j d_i - s_i d_j\}$$

where s_i is the integrated hazard divided by the 'fixed effect' function:

$$s_i = \frac{z_i}{g(w_i)} = \frac{\int_0^{c_i} \theta_i ds}{g(w_i)}.$$

Theorem 2A

Suppose Assumption Kernel, Assumption Bandwidth and Assumption Data hold. Define $Q(\gamma) = -g(\gamma)^2$ where $g(\gamma)$ is defined as above. If $g(\gamma)$ is continuous in γ and if either assumption I or II are satisfied,

I. $g(\gamma)$ is monotonic in γ and $E\sup_\gamma \|g(\gamma)\| < \infty$;

II. $\gamma \in \Theta$ which is compact; $Eg(\gamma) = 0$ only if $\gamma = \gamma_0$ and $E\sup_{\gamma\in\Theta} \|g(\gamma)\| < \infty$; then $\hat{\gamma} = \arg\max_\gamma Q_n(\gamma)$ is a consistent estimate for γ_0; i.e.

$$\hat{\gamma} = \arg\max_\gamma Q_n(\gamma) \to_p \gamma_0.$$

Proof:

The function $g(\gamma)$ is continuous in γ and it convergence to the objective functions of Theorem 2. To be more specific: Let $Eg_{2l}(\beta)$ for $l = 1, \ldots K$, denote the expectation of the estimating function of theorem 2 and $Q_{2;0} = -\sum_l g_{2l}^2(\beta)^2$ denote the objective function of theorem 2.

If assumption I then is converges in probability to is satisfied then $g(\gamma)$ is monotonic and $Q(\gamma) = -g(\gamma)^2$ is concave in γ and therefore $EQ(\gamma) = Q_{2;0}$ is uniquely maximized at the truth. Therefore, maximizing $Q_n(\gamma)$ gives a consistent estimate; see Newey and McFadden, 1994, Theorem 2.7.

If assumption II is satisfied then the conditions of Newey and McFadden, 1994, Theorem 2.6 are satisfied and $g(\beta) \to^U Eg_{2l}(\beta)$ and $Q(\beta) \to^U Q_{2;0}$. The objective function $Q_{2;0}$ is maximized at the truth consistency of $\hat{\gamma}$ follows. Q.E.D.

In theorem 1A and theorem 2A we assume that the "constructed fixed effect" depends only on the observable characteristics w_i, $v(w_i)$. We can relax this assumption a bit for uncensored spells. Let the 'constructed fixed effect' depend on the w_i and an error term, ε_i, that is independent of w_i and x_i: $v(w_i;\varepsilon_i)$. Or the "constructed fixed effect" can depend on unobserved characteristics, u_i, as well, i.e. $v(w_i;\varepsilon_i;u_i)$ and we want to estimate a single parameter model. The function $v(w_i;\varepsilon_i;u_i)$ is continuous in w_i and w_i is continuously distributed.

Theorem 2B

Suppose we observe N uncensored spells and want to estimate the exponential hazard model. Assume we are willing to match observations using the vector of regressors w. Define $g(\beta)$ as above and further assume that, for matched spells, $E(\frac{1}{v(w_j;\varepsilon_j)}) = E(\frac{1}{v(w_i;\varepsilon_i)})$ and $\varepsilon_i - \varepsilon_j$ are independent of $x_i - x_j$ so $Cov((x_i - x_j),(\frac{1}{v(w_i;\varepsilon_i)} - \frac{1}{v(w_j;\varepsilon_j)})) = 0$. Assume $\beta \in \Theta$, which is compact. Then estimate for β is consistent.

In many applications, $\frac{1}{v(w_j;\varepsilon_j)} - \frac{1}{v(w_i;\varepsilon_i)}$ is correlated with at least one element of x. In that case we can adjust theorem 2A and allow for that correlation. However, the matching becomes trickier since we want to assure monotonicity of $g(\beta)$: If $x_j < x_i$ for most pairs then $g(\beta)$ is monotonic and consistent inference is possible.

Theorem 2C

Suppose we observe N uncensored spells and want to estimate the single parameter exponential hazard model. Assume we are willing to match observations using the vector of regressors w: Define $g(\beta) = \binom{n}{2}^{-1} \sum_{j<i} \frac{1}{h_n^L} K(\frac{w_i - w_j}{h_n}) \{s_j d_i - s_i d_j\}$: Assume that, for matched spells, $E(\frac{1}{v(w_j;\varepsilon_j;u_j)}) = E(\frac{1}{v(w_i;\varepsilon_i;u_j)})$ and that $x_j < x_i$ for a sufficient number of pairs so that $g(\beta)$ is monotonic. The error term u may be correlated with x; i.e. the difference $\frac{1}{v_j} - \frac{1}{v_i}$ may be correlated with x but is on average zero. Assume $\beta \in \Theta$; which is compact. Then the estimator $\hat{\beta} = \arg\max -g(\beta)^2$ is a consistent estimate for β.

Note that theorem 2C does not restrict the relation between x and v to a particular form. This makes it a unique matching estimator[21]. Its result is driven by the fact that the matching is on average right and that the first spell has a value for x that is usually larger than that of the second spell.

The final case were we can use the moment functions of 1A and 2A is where u_i is only correlated with one regressor but independent of the other. The number of regressors is not restricted and theorem 2A replaces the estimating function for the dependent regressor in theorem 1A. The resulting method of moments estimator gives consistent estimates.

5.1 Concluding remarks about estimation using single spell data

Mathematically the "constructed fixed effect", $f_i = v(w_i)$; can be treated in nearly the same way as the fixed effects of the previous sections. However the motivation and assumptions on the data are different. The following list shows under which assumptions we can make consistent inference about the common parameters and it shows the difference between the assumptions for the 'constructed fixed effect' and the usual fixed effect of the previous sections.

1. The 'constructed fixed effect' depends only on the observable characteristics w_i: $v(w_i)$: The function $v(w_i)$ is continuous and w_i is continuously distributed.

2. The 'constructed fixed effect' depends on the w_i and an error term, ε_i; that is independent of w_i and x_i : $v(w_i;\varepsilon_i)$: The function $v(w_i;\varepsilon_i)$ is continuous in w_i and w_i is continuously distributed.

3a. The 'constructed fixed effect' can depend on unobserved characteristics, u_i; as well, i.e. $v(w_i;\varepsilon_i;u_i)$ and we want to estimate a single parameter model. The function $v(w_i;\varepsilon_i;u_i)$ is continuous in w_i and w_i is continuously distributed.

3b. As assumption 3a but now the u_i is only correlated with one regressor and the number of regressors is not restricted.

4. The fixed effect, f_i; can have any finite value which can differ for each i; i.e. no assumptions about the distributions of the fixed effects are made. This includes that f_i =

[21] The most common restriction being v and x being independent.

$v_i(w_i; \varepsilon_i; u_i)$ can differ across individuals, be discontinuous and that $\varepsilon_i; u_i$ can be correlated with x_i. Multiple observations are needed for each individual.

PH use assumption 1 and discuss several non-linear models including selection models. We will use assumption 1 trough 3b and focus on the moment conditions that were developed in the previous sections.

Mundlak (1961) argues that the fixed effect will often be correlated with the regressors, as when the regressor (input) is chosen in by an agent who knows his fixed effect f_i. If one knows which characteristics are important for the constructed fixed effect and if there is only one input then the assumptions of case 3a are satisfied. If there is more than one variable that is correlated with the fixed effect then none of the assumptions 1 through 3b hold. For linear models, Mundlak (1978) suggests to replace the fixed effect by a function that is a linear function of the regressors and a normally distributed error term, i.e. $f_i(x_i; e; \alpha) = x_i'\alpha + e$ where $e \sim N(0; \sigma^2)$. For a linear model we can estimate the common parameters by general least squares. By taking the logarithm of the first and second duration we can use that framework for the estimation of duration models[22] However, the linearity assumption seems often even less attractive than the matching assumptions of (1) through (3b).

It can be argued that allowing for fixed effects gives us a semiparametric hazard since the distribution of the individual effect given the regressor is not specified. That it would be a nice kind of semiparametric and in the case we have more than one observation per individual it is preferable above the semiparametric matching techniques that are discussed in this section.

6 Compare estimators

Cox' Estimator applied to a stratum

The partial likelihood estimator of Cox (1972) can be applied to different observations for the same individual. Chamberlain (1985) shows that the resulting estimator is consistent in the case that the durations are uncensored and the regressors are exogenous. In the framework of uncensored observations, Yamuguchi (1986) compares this estimator to random effects estimators and estimators for the exponential hazard model with fixed effects. Recently, the estimator is discussed by Ridder and Tunali (1999) and Lancaster (1999a, 6.8.2). The estimator that is consistent for the exponential hazard model as well as for other baseline hazards. That is, the hazard is

$$\theta_{is}(t) = e^{x_{is}\beta + v_i} k_i(t)$$

where $k_i(t)$ denotes an individual, unspecified baseline hazard. We can consider the rank ordering of $y_{i1}; y_{i2}$. If complete spells are observed then $\Pr(y_{i1} < y_{i2} | x_{i1}; x_{i2}; v_i; k_{i_0}(.)) = \frac{e^{x_{i1}}}{e^{x_{i1}} + e^{x_{i2}}}$. Lancaster (1999a) suggests to use the same formula for data with censored observations. If the censoring times are unequal for the different spells we can censor each individual at the smallest observed censoring time. Individuals for whom both spells are

[22] In the case that no spell is censored.

censored are discarded. Since an endogenous selection criterion is used (the outcome of the dependent variable determines whether an individual is discarded or not) consistency is not obvious. We discuss this estimator and then compare the asymptotic variance with the earlier estimator of theorem 1.

We can write the probability that the first spell lasts shorter than the second one as follows: $\Pr(y_{i1} < y_{i2} | x_{i1}; x_{i2}; v_i; k_{i_0}(.)) = \frac{e^{x_{i1}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}}$ gives us the following log likelihood:

$$L(\beta) = m x_{i1}\beta + (1 - m) x_{i2}\beta - \ln(e^{x_{i1}\beta} + e^{x_{i2}\beta})$$

where $m = 1(y_{i1} < y_{i2})$ and the subscript i is suppressed.

$$\begin{aligned} L_\beta &= m x_{i1} + (1 - m) x_{i2} - \frac{x_{i1} e^{x_{i1}\beta} + x_{i2} e^{x_{i2}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} \\ &= m(x_{i1} - x_{i2}) + x_{i2} - \frac{x_{i1} e^{x_{i1}\beta} + x_{i2} e^{x_{i2}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} \end{aligned}$$

We introduce artificial censoring by the econometrician at the minimum censoring time of the individual: So the econometrician censors at $c = \min_s(c_{is})$ and discards observations that are censored twice. This endogenous selection does not affect the consistency of the estimator:

For duration models with hazard $\theta_{is}(y) = e^{x_{is}\beta + v_i} k_i(y)$ with two spells for each individual and exogenous censoring we can write the log likelihood of the order statistic as

$$L_i(\beta) = m x_{i1}\beta + (1 - m) x_{i2}\beta - \ln(e^{x_{i1}\beta} + e^{x_{i2}\beta}).$$

Consider the following estimator:

$$\hat{\beta} = \arg\max \frac{\sum_i L_i(\beta)}{N}$$

$\hat{\beta}$ is a consistent estimate for β_0. See Appendix 14 for a consistency proof.

Asymptotic variance

The asymptotic variance for the rank estimator is (see Appendix 15 for details):

$$\begin{aligned} \operatorname{var} L_\beta &= \frac{\sum_{i=1}^N (x_1 - x_2)^2 \operatorname{var}(m)}{N^2} \\ &= \frac{\sum_{i=1}^N (x_1 - x_2)^2 (\frac{1}{e^{x_{i1}\beta} + e^{x_{i2}\beta}})^2}{N^2} \quad \text{since } \Pr(m = 0) = \frac{e^{x_{i1}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} \\ &= \frac{\sum_{i=1}^N (\frac{x_1 - x_2}{e^{x_{i1}\beta} + e^{x_{i2}\beta}})^2}{N^2} \end{aligned}$$

Note that this is also the expression of the asymptotic variance in the case that the data are not censored. However, in the case of censoring we discard observations, so N is smaller than when there would be no censored observations. Assume that a spell will be censored if $t_{is} > c$. If the second observations have a chance of p of being censored at censoring time c

(i.e. $\Pr(t_{i2} > c) = p$) then the chance that the data of that individual are being discarded are:

$$
\begin{aligned}
\Pr(i \text{ discarded}) &= \Pr(\{t_{i1} > c \text{ and } t_{i2} > c) \\
&= \Pr(t_{i1} > c)\Pr(t_{i2} > c) \\
&= \frac{e^{x_{i2}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} p \\
&= \frac{e^{x_{i2}}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} p.
\end{aligned}
$$

So the fraction of observation that we use is:

$$
\begin{aligned}
\Pr(i \text{ not discarded}) &= 1 - \Pr(i \text{ discarded}) \\
&= 1 - \frac{e^{x_{i2}}}{e^{x_{i1}} + e^{x_{i2}}} p.
\end{aligned}
$$

6.1 Comparison of the asymptotic variances

Now we have two estimators for the exponential hazard model. Since both are consistent we can compare them by comparing their asymptotic variances for the case $T = 2$. For the first estimator, $Q_\beta(\beta) = (x_{i1} - x_{i2})\{e^{x_{i2}\beta} y_{i2} d_{i1} - e^{x_{i1}\beta} y_{i1} d_{i2}\}$ the information equality does not hold.

The asymptotic variance, σ^2_β is (see appendix 14)

$$
\begin{aligned}
\sigma^2_\beta &= J^{-1} I J^{-1} \\
&= \frac{(x_{i1} - x_{i2})^2}{2} \left\{ \frac{1}{\Pr(d_{i1} = 1)\Pr(d_{i2} = 1)} \right\}^2 \\
&\quad \left\{ 1 + \frac{z_{i1\max} z_{i2\max} e^{-(z_{i1\max} + z_{i2\max})}}{(1 - e^{-z_{i1\max}})(1 - e^{-z_{i2\max}})} \right\} \\
&= \frac{(x_{i1} - x_{i2})^2}{2} \left\{ \frac{1}{(1 - e^{-z_{i1\max}})(1 - e^{-z_{i2\max}})} \right\}^2 \\
&\quad \left\{ 1 + \frac{z_{i1\max} z_{i2\max} e^{-(z_{i1\max} + z_{i2\max})}}{(1 - e^{-z_{i1\max}})(1 - e^{-z_{i2\max}})} \right\}.
\end{aligned}
$$

The rank estimator has the following asymptotic variance

$$
\sigma^2_\beta = \frac{N}{\operatorname{var} L_\beta} = \left(\frac{\sum_{i=1}^{N} \left(\frac{x_1 - x_2}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} \right)^2}{N} \right)^{-1}.
$$

7 Conclusion

In this paper we use the integrated hazard to derive estimators for duration models. Van den Berg notes in his handbook chapter[23] that current estimation methods of duration models

[23] "Duration Models: Specification, Identification, and Multiple Duration," in Handbook of Econometrics, Vol. 5. Amsterdam: North-Holland.

are sensitive to censoring. For a wide range of models we file that "caveat". We further derived consistent estimators for fixed effect models that can deal with

- time varying regressors (e.g. unemployment rate)
- lagged duration dependence
- regressors that are endogenous in the sense that they depend on the length of earlier spells.

We do not strongly belief in "matching models"[24] but showed consistency of the estimators. We summarize our findings by saying that the integrated hazard is a useful concept in the sense that it can suggest consistent estimators.

8 Appendices

Appendix 1

We define $z_{is;max} = \int_0^{c_{is}} \mu(s;x)ds$

$$
\begin{aligned}
Ed_{is} &= \Pr\{t_{is} \leq c_{is}\} = \Pr\left\{\int_0^{t_{is}} \mu(s;x)ds \leq \int_0^{c_{is}} \mu(s;x)ds\right\} = \Pr\left\{z_{is} \leq \int_0^{c_{is}} \mu(s;x)ds\right\} \\
&= 1 - e^{-z_{is;max}}
\end{aligned}
$$

$$
\begin{aligned}
Ez_{is} &= \int_0^{z_{is;max}} z_{is}e^{z_{is}}dz_{is} + z_{z_{is;max}}e^{z_{z_{is;max}}} \Pr(z_{is} = z_{is;max}) \\
&= 1 - e^{-z_{is;max}} = Ed_{is}.
\end{aligned}
$$

Q.E.D.

Appendix 2

Proof of theorem 1:

The moments function has expectation zero at the truth, i.e. $Eg(\beta_0) = 0$ and therefore $Q_0(\beta_0) = -\{Eg(\beta_0)\}'\{Eg(\beta_0)\} = 0$; i.e. $Q_0(\beta_0)$ is maximized at the truth. The difficult part of the proof is to show that $Q_0(\beta_0)$ is a unique maximum. A sufficient condition for this is that $Eg(\beta) = 0$; is uniquely solved for $\beta = \beta_0$. In the case that β is a parameter and we observe two durations for each individual (i.e. $K = 1$ and $T = 2$) then the objective function $Q_n(\beta)$ is concave. We first give the proof for this easy case and than proof that $Q_0(\beta_0)$ has a unique solution in the general case. For $K = 1$ and $T = 2$ we have:

$$
\begin{aligned}
\mu_{i1} &= v_i \\
\mu_{i2} &= v_i e^{\Delta x_i \beta}
\end{aligned}
$$

[24] Since individuals differ in more ways than is revealed by the regressors.

which gives

$$g(\beta) = -\frac{\sum_{i=1}^{N} \Delta x_i \{e^{\Delta x_i \beta} y_{i2} d_{i1} - y_{i1} d_{i2}\}}{N}$$

and

$$\frac{\partial g(\beta)}{\partial \beta} = -\frac{\sum_{i=1}^{N} (\Delta x_i)^2 \{e^{\Delta x_i \beta} y_{i2} d_{i1}\}}{N} < 0 \text{ for all } \beta.$$

Since the derivative is always negative the solution to $g(\beta) = 0$ is unique. Therefore $Eg(\beta_0) = 0$ is uniquely solved for $\beta = \beta_0$.

For the general case we need that $Eg(\beta) = 0$ is uniquely solved for $\beta = \beta_0$, where

$$g(\beta) = \frac{1}{N} \sum_i g_i(\beta)$$

and

$$g_i(\beta) = \sum_{s<r} f(x_{ir} - x_{is})(d_{ir} e^{x_{is}\beta} y_{is} - d_{is} e^{x_{ir}\beta} y_{ir})\} \quad s, r = 1, \ldots, T_i.$$

Note that β and $(x_{ir} - x_{is})$ are vectors with length K (the number of regressors). It was shown in section 2.2.1 that

$$Eg(\beta_0) = 0.$$

We define $k_{is} = E(d_{is}) = Ee^{x_{is}\beta_0} y_{is}/f_i$, $g_{is} = e^{x_{is}(\beta - \beta_0)}$ and a weighted mean of g_i for each individual, $\tilde{g}_i = \frac{1}{T_i} \sum_s g_{is} k_{is} = \frac{1}{T_i} \sum_s e^{x_{is}(\beta-\beta_0)} k_{is}$. The following relations hold:

$$\begin{aligned}
Eg(\beta) &= \frac{1}{N} \sum_i E[\sum_{s<r} f(x_{ir} - x_{is})(d_{ir} e^{x_{is}\beta} y_{is} - d_{is} e^{x_{ir}\beta} y_{ir})\}] \\
&= \frac{1}{N f_i} \sum_i [\sum_{s<r} f(x_{ir} - x_{is})(k_{ir} g_{is} k_{is} - k_{is} g_{ir} k_{ir})\}] \\
&= \frac{1}{N f_i} \sum_i [\sum_{s<r} f(x_{ir} - x_{is})(g_{ir} - g_{is}) k_{ir} k_{is}\}].
\end{aligned}$$

If $\beta = \beta_0$ then $g_{ir} = g_{is} = 1$ and $g_{ir} - g_{is} = 0$ so $Eg(\beta_0) = 0$. If $g_{ir} - g_{is}$ depends on $(x_{ir} - x_{is})$ (e.g. if $\beta \neq \beta_0$) then it seems that $g_{ir} - g_{is}$ is 'correlated' with $(x_{ir} - x_{is}) k_{ir} k_{is}$ and $Eg(\beta) \neq 0$. Therefore the only solution seems to be that g_{is} is a constant. We will now argue that g_{is} is only a constant at the true value β_0. The function g_{is} was defined as $g_{is} = e^{x_{is}(\beta-\beta_0)}$; the matrix X has full rank and therefore there is no $\beta' \neq \beta_0$ s.t. $x_{is}(\beta' - \beta_0) = 0$; and therefore there is no $\beta' \neq \beta_0$ s.t. $g_{is} = e^{x_{is}(\beta-\beta_0)}$ is a constant (i.e. the only value for which g_{is} does not vary with x_{is} is β_0). Therefore, $\beta = \beta_0$ is the unique solution of $Eg(\beta) = 0$.

We use lemma 2.4 of Newey and McFadden to prove the conditions for uniform convergence: The data are i.i.d. (follows from the hazard function), the parameter space is compact (was assumed), $g_l(\pi;\gamma)$ is continuous and has finite expectation (dominance condition) for all β and all l. Using lemma 2.4 uniform convergence is ensured.

The limit of the objective function, $Q_0(\beta)$ is uniquely maximized at the truth. The objective function $Q_n(\beta) = g^2(\beta)$ converges uniformly to $Q_0(\beta)$. Therefore, maximizing $Q(\beta)$ gives a consistent estimate (see Newey and McFadden, 1994, page 2133). Q.E.D.

Appendix 3

Proof of lemma 2:

$$\begin{aligned} Eg(\gamma_0) &= E\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\{\sum_{s<r}(d_{ir}z_{is} - d_{is}z_{ir})\Big\}\Big] \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big\{\sum_{s<r}(Ed_{ir}Ez_{is} - Ed_{is}Ez_{ir})\Big\} \text{ using independence.} \end{aligned}$$

By lemma 1 we have $Ez_{is} = E\int_0^{y_{is}}\mu(s;x)ds = Ed_{is}$ and therefore

$$Eg(\gamma_0) = \frac{1}{N}\sum_{i=1}^{N}\Big\{\sum_{s<r} Ed_{is}Ed_{ir} - Ed_{ir}Ed_{is}\Big\} = 0.$$

Q.E.D.

Appendix 4

$$g^*(\beta;\gamma) = \begin{pmatrix} g_W(\beta;\gamma) \\ g_E(\beta;\gamma) \end{pmatrix}$$

Note that

$$\begin{aligned} Eg_W(\beta_0;\gamma_0) &= \frac{1}{N}\sum_{i=1}^{N}\Big\{\sum_{s<r} E(d_{ir}e^{x_{is}\beta}y_{is}^{\gamma} - d_{is}e^{x_{ir}\beta}y_{ir}^{\gamma})\Big\} \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big\{\sum_{s<r}(Ed_{ir}E\varsigma_{is} - Ed_{is}E\varsigma_{ir})\Big\} \\ &= 0 \text{ since } Ed_{ir} = E\varsigma_{ir} \text{ and } E\varsigma_{is} = Ed_{is} \end{aligned}$$

and

$$\begin{aligned} Eg_E(\beta_0;\gamma_0) &= \frac{1}{N}\sum_{i=1}^{N}\Big[\sum_{s<r}\{(x_{ir} - x_{is})E(d_{ir}e^{x_{is}\beta}y_{is}^{\gamma} - d_{is}e^{x_{ir}\beta}y_{ir}^{\gamma})\}\Big] \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big[\sum_{s<r}\{(x_{ir} - x_{is})(Ed_{ir}Ez_{is} - Ed_{is}Ez_{ir})\}\Big] \\ &= 0 \text{ since } Ed_{ir} = Ez_{ir} \text{ and } Ed_{ir} = Ez_{ir}. \end{aligned}$$

The rest of the proof is similar to the proof of theorem 1; the extra moment condition requires that $E(k_{ir}g_{is}k_{is} - k_{is}g_{ir}k_{ir}) = 0$.

Appendix 5

Proof of theorem 4: Note that L_n, the number of individuals in each interval, and M, the number of intervals, are defined in such a way that as $N \to \infty$ then $L_n, M \to \infty$. $Q_0 = -\sum_{l=1}^{M} \{g_{0,l}(\gamma)\}^2$ is uniquely maximized at the truth (see above).

We use lemma 2.4 of Newey and McFadden to prove the conditions for uniform convergence: The data are i.i.d. (follows from the hazard function), the parameter space is compact (was assumed), $g_l(\gamma)$ is continuous and bounded in probability for all γ and all l. Using lemma 2.4 uniform convergence is ensured so $g_l(\gamma) \xrightarrow{U} g_{0,l}(\gamma)$ and therefore $Q_n(\gamma) \xrightarrow{U} Q_0$.

We write the mean square error (MSE) as the addition of the variance and the squared bias:

$$\begin{aligned} MSE &= E\{(\hat{\lambda}_n(t) - \lambda_0(t))^2\} \\ &= E\{(\hat{\lambda}_n(t) - \mu + \mu - \lambda_0(t))^2\} \text{ where } \mu = E(\hat{\lambda}_n(t)) \\ &= E\{(\hat{\lambda}_n(t) - \mu)^2\} + E\{(\mu - \lambda_0(t))^2\}. \end{aligned}$$

The limit of the variance is zero for $n, k \to \infty$:

$$\lim_{n,k \to \infty} E\{(\hat{\lambda}_n(t) - \mu)^2\} = 0.$$

The limit of the bias is zero as well as $n, k \to \infty$.
Therefore, the limit of the MSE is zero.

$$\lim_{n,k \to \infty} E\{(\mu - \lambda_0(t))^2\} = 0.$$

Q.E.D.

Appendix 6

Proof of theorem 5: Similar to the proof of theorem 4. The limit of the objection function, $Q_0 = -\{\sum_{l=1}^{M} \{g_{0,l}(\gamma)\}^2 + \sum_{r=1}^{K} \{g_{0,r}(\gamma)\}^2\}$ is uniquely maximized at the truth, $g_l(\gamma)$ converges uniformly to $g_{0,l}(\gamma)$ and $Q_n(\gamma)$ converges uniformly to Q_0 and $\lim_{n \to \infty} E\{(\hat{\lambda}_n(t) - \lambda_0(t))^2\} = 0$.Q.E.D.

Appendix 7

$$\tilde{d}_{i2} = d_{i2}^{Data} + (1 - d_{i2}^{Data}) ind$$

So

$$E\tilde{d}_{i2} = Ed_{i2}^{Data} + E\{(1 - d_{i2}^{Data}) ind\}.$$

As was explained in the text, $Ed_{i2}^{Data} = E(d_{i2} | t_{i2} \le c_2^{Data}) P(t_{i2} \le c_2^{Data})$ and $E\{(1 - d_{i2}^{Data}) ind\} = (d_{i2} | t_{i2} > c_2^{Data}) P(t_{i2} > c_2^{Data})$. Therefore,

$$E\tilde{d}_{i2} = E(d_{i2} | t_{i2} \le c_2^{Data}) P(t_{i2} \le c_2^{Data}) + E(d_{i2} | t_{i2} > c_2^{Data}) P(t_{i2} > c_2^{Data})$$

where the last expression is, by definition, equal to Ed_{i2}.

Appendix 8

Proof of lemma 4:

$$
\begin{aligned}
E(z_{is}d_{is}) &= E(z_{is}|d_{is}=1)\Pr(d_{is}=1) \\
&= \int_0^{z_{is\max}} z\frac{e^{-z}}{1-e^{-z_{is\max}}}dz(1-e^{-z_{is\max}}) \\
&= \int_0^{z_{is\max}} ze^{-z}dz \\
&= 1-z_{is\max}e^{-z_{is\max}}-e^{-z_{is\max}}.
\end{aligned}
$$

$$
\begin{aligned}
Eg(\gamma_0) &= \frac{1}{f_i}E(z_{i1}d_{i1}d_{i2}-z_{i2}d_{i2}d_{i1}) \\
&= \frac{1}{f_i}E(z_{i1}d_{i1}d_{i2}-z_{i2}d_{i2}d_{i1}) \\
&= \frac{1}{f_i}E(z_{i1}d_{i1}|d_{i2}=1)\}\Pr(d_{i2}=1)-\frac{1}{f_i}\{E(z_{i2}d_{i2}|d_{i1}=1)\}\Pr(d_{i1}=1).
\end{aligned}
$$

Where $\Pr(d_{i2}=1)=\Pr(d_{i2}=1)=1-e^{-z_{max}}$. The conditional expectations, $E(z_{i1}d_{i1}|d_{i2}=1)$ and $E(z_{i2}d_{i2}|d_{i1}=1)$ equal the unconditional expectations, i.e. $E(z_{i1}d_{i1}|d_{i2}=1)=E(z_{i1}d_{i1})$ and $E(z_{i2}d_{i2}|d_{i1}=1)=E(z_{i2}d_{i2})$ and therefore

$$
\begin{aligned}
Eg(\gamma_0) &= \frac{1}{f_i}\{E(z_{i1}d_{i1})-E(z_{i2}d_{i2})\}(1-e^{-z_{max}}) \\
&= \frac{1}{f_i}\{E(z_{i1}d_{i1})-E(z_{i2}d_{i2})\}(1-e^{-z_{max}}) \text{ using Lemma 3} \\
&= 0 \text{ since } z_{i1\max}=z_{i2\max} \text{ and therefore } E(z_{i1}d_{i1})=E(z_{i2}d_{i2}).
\end{aligned}
$$

Q.E.D.

Appendix 9

Proof of theorem 6:

The standard conditions for $Q(\beta)$ hold:

(i) $Eg(\beta_0)=0$ (see above);

(ii) $g(\beta)$ is continuous in β (by inspection);

(iii) $\frac{\partial Eg(\beta)}{\partial\beta}<0$ $g(\beta)$ is monotonic in all elements of β since their derivatives are always negative (see above); {complete the proof that $Eg(\beta)=0$ is uniquely solved for $\beta=\beta_0$.};

(iv) $EQ(\beta)=-E\{g(\beta)'Ig(\beta)\}$ is concave and therefore $EQ(\beta)$ is uniquely maximized at the truth, β_0;

(v) Q converges uniformly to EQ {to be proven, seems to hold}.

Therefore, maximizing $Q(\beta)$ gives a consistent estimate for β_0 (see Newey and McFadden, 1994, page 2133). Q.E.D.

Appendix 10

The standard conditions for concave extremum estimators hold:

(i) $Eg(\gamma_0) = 0$ by Lemma 4;

(ii) $g(\gamma)$ is continuous in β (by inspection);

(iii) $g_\gamma(\gamma) = -a(t_1)t_{i2}d_{i2}d_{i1} < 0$ so $g_\gamma(\gamma)$ is monotonic in γ and $Q(\gamma) = -(\frac{\sum g_\gamma(\gamma)}{\sum d_{i1}d_{i2}})^2$ is concave in γ. So $EQ(\gamma)$ is uniquely maximized at the truth;

(iv) $g(\gamma)$ converges in probability to $E\,g(\gamma)$ (no need for uniform convergence since $Q(\gamma)$ is concave in γ.

Therefore, maximizing $Q(\gamma)$ gives a consistent estimate for γ_0 (see Newey and McFadden, 1994, page 2133). Q.E.D.

Appendix 11

To be proven:

$$Eg^*(\gamma_0) = \frac{1}{N}E\sum_i \bar{w}_i^m\, g_i(\gamma_0) = 0$$

where $\alpha_c^m \frac{1}{f_i}$

$$g_i(\gamma_0) = t_{i1} - e^{\Delta w_i \gamma} t_{i2}.$$

Note that

$\frac{1}{N}\sum_i \bar{w}_i^m t_{i1} = \frac{1}{N}\sum_i (\alpha_c^m + e_i^m)t_{i1} = \frac{1}{N}\alpha_c^m \sum_i t_{i1} - \sum_i e_i^m t_{i1} = \alpha_c^m \bar{t}_1$ since $\sum_i e_i^m t_{i1} = 0$.

Using the law of iterated expectations: $E(\alpha_c^m \bar{t}_1) = E\{E(\alpha_c^m \bar{t}_1 | t_{i1}, \ldots, t_{iN})\} = \alpha_c^m E\bar{t}_1 = \alpha_c^m \sum_i \frac{1}{N f_i}$

$E\bar{w}_i^m e^{w_{is}\gamma_0} t_{i2} = E(\alpha_c^m + e_i^m) e^{w_{is}\gamma_0} t_{i2} = E\{(\alpha_c^m + e_i^m)\frac{z_{i2}}{f_i}\} = E\alpha_c^m \frac{z_{i2}}{f_i} = \alpha_c^m \sum_i \frac{1}{N f_i}$.

Therefore, $Eg^*(\gamma_0) = 0$.

Appendix 12

Proof of theorem 8.

Define $\bar{w}_i = \{\bar{w}_i^1; \bar{w}_i^2; \ldots; \bar{w}_i^M\}$. The matrix with rows $\{t_{i1}; \Delta w_i'\}$ has full column rank and therefore the matrix with rows $\{t_{i1}; \bar{w}_i^1; \bar{w}_i^2; \ldots; \bar{w}_i^M\} = \{t_{i1}; \bar{w}_i\}$ has full column rank as well. Consider estimating the following hazard:

$$\begin{aligned} \theta_{i1} &= f_i \\ \theta_{i2} &= f_i e^{\Delta \bar{w}_i \gamma + \delta t_{i1}}. \end{aligned}$$

Theorem 1 and 2 suggest the following moment equations:

$$h(\gamma; \delta) = \frac{1}{N}\sum_i (t_{i1} - e^{\Delta \bar{w}_i \gamma + \delta t_{i1}} t_{i2})$$

$$h^*(\gamma; \delta) = \frac{1}{N}\sum_i \{\Delta \bar{w}_i (t_{i1} - e^{\Delta \bar{w}_i \gamma + \delta t_{i1}} t_{i2})\}$$

and consistent estimation of $\{\gamma, \delta\}$ is possible:

(i) $Eh(\gamma_0; \delta_0) = Eh^*(\gamma_0; \delta_0) = 0$, arguments similar to $Eg(\gamma_0) = Eg^*(\gamma_0) = 0$ in the text;
(ii) $h^*(\gamma; \delta)$ is monotonic decreasing in each element of γ since the derivative is always negative:

$$\frac{\partial h^*(\gamma; \delta)}{\partial \gamma_l} = \frac{\sum_i -(\Delta w_i)^2 e^{\Delta w_i \gamma + \delta t_{i1}}) t_{i2}}{N} < 0;$$

(iii) $h(\gamma; \delta)$ is monotonic decreasing in δ since the derivative is always negative:

$$\frac{\partial h(\gamma; \delta)}{\partial \delta} = \frac{\sum_i -\Delta w_i e^{\Delta w_i \gamma + \delta t_{i1}}) t_{i1} t_{i2}}{N} < 0.$$

For a one parameter model it is enough that one parameter $g()$ is monotonic in its parameter to ensure uniqueness of $g() = 0$. For a model with more parameters we know that the matrix with rows $\{t_{i1}; w_i\}$ has full rank. So only at $\{\gamma; \delta\} = \{\gamma_0; \delta_0\}$ we have $\Delta w_i \gamma + \delta t_{i1} = 0$ for all i: If $\Delta w_i \gamma + \delta t_{i1} \neq 0$ then the value of this equation will be correlated with Δw_i and t_{i1}: The moment function $h^*(\gamma; \delta)$ "adds up" without weighting the observation; the other equations weight according to the values of Δw_i; so correlation causes at least one equation to be nonzero, see the proof of theorem 1, appendix 1. Therefore, the maximum of $Q(\gamma; \delta) = -\{h(\gamma; \delta)^2 + h^*(\gamma; \delta)' h^*(\gamma; \delta)\}$ is at the truth;

(iv) We use lemma 2.4 of Newey and McFadden to prove the conditions for uniform convergence: The data are i.i.d. (follows from the hazard function), the parameter space is compact (was assumed), $g_l(\pi; \gamma)$ is continuous and bounded in probability for all $\pi; \gamma$ and all l: Using lemma 2.4 uniform convergence is ensured.

Consider the restriction $\delta = \gamma' \alpha$ where $\alpha = \{\alpha_s^1; \alpha_s^2; \ldots; \alpha_s^M\}$; this restriction does not effect identification or consistency. Plugging this restriction in the hazard for the second period gives

$$\theta_{i2} = f_i e^{\Delta w_i \gamma + \delta t_{i1}} = f_i e^{\Delta w_i \gamma}$$

which is the hazard of theorem 8 and consistency of that hazard model follows.

Appendix 12 A

{prove concavity in Theorem 5}

$$g(\beta) = (x_{i2} - x_{i1})\{e^{x_{i1}\beta} y_{i1} d_{i1} d_{i2} - e^{x_{i2}\beta} y_{i2} d_{i2} d_{i2}\}$$

$$g(\beta) = d_{i1} d_{i2} (x_{i2} - x_{i1})\{e^{x_{i1}\beta} y_{i1} - e^{x_{i2}\beta} y_{i2}\}$$

$$g(\beta)^2 = (d_{i1} d_{i2})^2 (x_{i2} - x_{i1})^2 (e^{x_{i1}\beta} y_{i1} - e^{x_{i2}\beta} y_{i2})^2$$

$$\begin{aligned} \frac{\partial g(\beta)^2}{\partial \beta} &= 2(d_{i1} d_{i2})^2 (x_{i2} - x_{i1})^2 (e^{x_{i1}\beta} y_{i1} - e^{x_{i2}\beta} y_{i2})(x_{i1} e^{x_{i1}\beta} y_{i1} - x_{i2} e^{x_{i2}\beta} y_{i2}) \\ &= 2(d_{i1} d_{i2})^2 (x_{i2} - x_{i1})^2 \{x_{i1}(e^{x_{i1}\beta} y_{i1})^2 - (x_{i1} + x_{i2}) e^{x_{i1}\beta} y_{i1} e^{x_{i2}\beta} y_{i2} - x_{i2}(e^{x_{i2}\beta} y_{i2})^2\} \end{aligned}$$

$$\{x_{i1}(e^{x_{i1}\beta}y_{i1})^2 - (x_{i1} + x_{i2})e^{x_{i1}\beta}y_{i1}e^{x_{i2}\beta}y_{i2} - x_{i2}(e^{x_{i2}\beta}y_{i2})^2\}$$

$x_{i2} = -x_{i1}$

$$\{x_{i1}(e^{x_{i1}\beta}y_{i1})^2 + x_{i1}(e^{x_{i2}\beta}y_{i2})^2\}.$$

Appendix 12 B

At the truth, the expectation of the moment function equals zero.

$$Eg(\gamma_0) = \frac{1}{N f_i}\sum_i E\{d_{i2}y_{i1} - d_{i1}e^{\gamma r(y_i)}y_{i2}\} = \frac{1}{N f_i}\sum_i E\{d_{i2}z_{i1} - d_{i1}z_{i2}\} = 0.$$

For uncensored observations, $\frac{\partial g_i(\gamma)}{\partial \gamma} = -e^{\gamma r(y_i)}y_{i2} < 0.$ Note that $Ed_{i2} = 1 - e^{z_{1\max}}$ where $z_{1\max} = e^{-\gamma r(y_i)}e^{\gamma_0 r(y_i)}c_{i2}$ and $\frac{\partial Ed_{i2}}{\partial \gamma} > 0$. Furthermore, $Ey_{i2} = Ed_{i2}$ and the derivatives are equal as well and in general, $Eg(\gamma)$ is uniquely solved for $\gamma = \gamma_0.$

Appendix 13

Proof of theorem 9:

The moments function has expectation zero at the truth, i.e.

$$\begin{aligned} Eg^{PC}(\beta_0;\gamma_0) &= 0 \\ Eg^{E}(\beta_0;\gamma_0) &= 0 \\ Eg^{*E}(\beta_0;\gamma_0) &= 0. \end{aligned}$$

We can a construct the usual GMM objective functions. Giving equal weight to all the moments gives

$$Q_n(\beta;\gamma) = -\{g^{PC}(\beta;\gamma)\}'\{g^{PC}(\beta;\gamma)\} - \{g^{E}(\beta;\gamma)\}^2 - \{g^{*E}(\beta;\gamma)\}'\{g^{*E}(\beta;\gamma)\}$$

$$Q_0(\beta_0) = -\{Eg^{PC}(\beta_0;\gamma_0)\}'\{Eg^{PC}(\beta_0;\gamma_0)\} - \{Eg^{E}(\beta_0;\gamma_0)\}^2 - \{Eg^{*E}(\beta_0;\gamma_0)\}'\{Eg^{*E}(\beta_0;\gamma_0)\} = 0$$

Thus $Q_0(\beta_0)$ is maximized at the truth. Arguments similar to earlier proofs (theorem 1 and 8) show that $Q_0(\beta_0)$ is a unique maximum. We use lemma 2.4 of Newey and McFadden to prove the conditions for uniform convergence: The data are i.i.d. (follows from the hazard function), the parameter space is compact (was assumed), the moment functions are is continuous and has finite expectation (dominance condition) for all β and all $\gamma.$ Using lemma 2.4 uniform convergence is ensured.

The limit of the objective function, $Q_0(\beta;\gamma)$ is uniquely maximized at the truth. The objective function $Q_n(\beta;\gamma)$ converges uniformly to $Q_0(\beta;\gamma).$ Therefore, maximizing $Q(\beta;\gamma)$ gives a consistent estimate (see Newey and McFadden, 1994, page 2133). Q.E.D.

Appendix 14

Notation: $f_s = f(y_{is}) = \mu_{is}(y_{is})e^{-\int_0^{y_{is}}\mu_{is}(s)ds}$;

$\bar{F}_s(y_{is}) = \bar{F}(y_{is}) = e^{-\int_0^{y_{is}}\mu_{is}(s)ds}$; note that $\bar{F}_s(y_{i1}) = e^{-\int_0^{y_{is}}\mu_{i1}(s)ds} = e^{-\alpha\int_0^{y_{is}}\mu_{i2}(s)ds} = (\bar{F}_s(y_{i2}))^{\alpha}$ with $\alpha = e^{(x_{i1}-x_{i2})\beta}$

and $\bar{F}_s = \bar{F}_s(c)$ where $c = \min_s(c_{i1}, c_{i2})$

$$
\begin{aligned}
\Pr(t_{i1} &< t_{i2}|t_{i1}, t_{i2} < c) = 1 - \Pr(t_{i1} < t_{i2}|t_{i1}, t_{i2} > c) \\
&= \Pr(t_{i1} > t_{i2}|t_{i1}, t_{i2} > c).
\end{aligned}
$$

notation:

$$
\begin{aligned}
\Pr(t_{i1} &> t_{i2}|t_{i1}, t_{i2} > c) = \int_c^{\infty}\left\{\int_{t_2}^{\infty}\frac{f_1}{\bar{F}_1}dt_1\right\}\frac{f_2}{\bar{F}_2}dt_2 \\
&= \int_c^{\infty}\left\{\frac{1-F_1(t_2)}{\bar{F}_1}\right\}\frac{f_2}{\bar{F}_2}dt_2 \\
&= \int_c^{\infty}\left\{\frac{\bar{F}_1(t_2)}{\bar{F}_1}\right\}\frac{f_2(t_2)}{\bar{F}_2}dt_2 \\
&= \int_c^{\infty}\frac{f_2(t_2)}{\bar{F}_2}\left\{\frac{\bar{F}_2(t_2)}{\bar{F}_2}\right\}^{\alpha}dt_2 \text{ since } \bar{F}_1() = \bar{F}_2^{\alpha} \text{ where } \alpha = e^{(x_{i1}-x_{i2})\beta} \\
&= \left|-\frac{1}{\alpha+1}\left\{\frac{\bar{F}_1(t_2)}{\bar{F}_2}\right\}^{\alpha+1}\right|_c^{\infty} \\
&= \frac{1}{\alpha+1} = \frac{1}{e^{(x_1-x_2)\beta}+1} \\
&= \frac{e^{x_2\beta}}{e^{x_1\beta}+e^{x_2\beta}}.
\end{aligned}
$$

And therefore

$$
\Pr(t_{i1} < t_{i2}|t_{i1}, t_{i2} > c) = \frac{e^{x_1\beta}}{e^{x_1\beta}+e^{x_2\beta}}.
$$

So $\Pr(t_{i1} < t_{i2}|t_{i1}, t_{i2} > c) = \Pr(t_{i1} < t_{i2})$.

Now we can calculate $\Pr(t_{i1} < t_{i2}|t_{i1}, t_{i2} < c)$:

$$
\begin{aligned}
\Pr(t_{i1} &< t_{i2}) = \Pr(t_{i1} < t_{i2}|t_{i1}, t_{i2} < c)P(t_{i1}, t_{i2} < c) \\
+\Pr(t_{i1} &< t_{i2}|t_{i1}, t_{i2} > c)P(t_{i1}, t_{i2} > c).
\end{aligned}
$$

And therefore:

$$
\begin{aligned}
\Pr(t_{i1} &< t_{i2}|t_{i1}, t_{i2} < c) = \frac{\Pr(t_{i1} < t_{i2})}{P(t_{i1}, t_{i2} < c)}\{1 - P(t_{i1}, t_{i2} > c)\} \\
&= \Pr(t_{i1} < t_{i2}).
\end{aligned}
$$

So the rank ordering is correct and indeed is the expectation of the score zero:

$$
E[m|t_{i1}, t_{i2} < c_i] = \Pr(t_{i1} < t_{i2}|t_{i1}, t_{i2} < c) = \frac{e^{x_1\beta}}{e^{x_1\beta}+e^{x_2\beta}}.
$$

And therefore:

$$
\begin{aligned}
E\dot{L}|_{\beta=\beta_0} &= (x_1 - x_2)Em + x_2 - \frac{x_{i1}e^{x_{i1}\beta} + x_{i2}e^{x_{i2}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} \\
&= (x_1 - x_2)\frac{e^{x_1\beta}}{e^{x_1\beta} + e^{x_2\beta}} + x_2 - \frac{x_{i1}e^{x_{i1}\beta} + x_{i2}e^{x_{i2}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} \\
&= -x_2\frac{e^{x_1\beta}}{e^{x_1\beta} + e^{x_2\beta}} + x_2 - \frac{x_{i2}e^{x_{i2}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}} = 0.
\end{aligned}
$$

$L(\beta)$ is concave in β and consistency follows. Q.E.D.

Appendix 15

By the usual delta method, we expand $\dot{Q}(\beta)$ around $\hat{\beta}$ and assume that the law of large numbers applies:

$$
\sqrt{n}(\hat{\beta} - \beta_0) = \frac{N}{\ddot{Q}(\beta_0)}\frac{\dot{Q}(\beta_0)}{\sqrt{n}}.
$$

The Information matrix and the Hessian are $I = \operatorname{var}\dot{Q}(\beta_0)$
$J = E\ddot{Q}(\beta_0)$ and therefore

$$
\sqrt{n}(\hat{\beta} - \beta_0) \sim N(0; J^{-1} I J^{-1}).
$$

Calculation of the Hessian:

$$
\begin{aligned}
J &= E\ddot{Q}(\beta_0) = E(x_{i1} - x_{i2})\{x_{i2}e^{x_{i2}\beta_0}y_{i2}d_{i1} - x_{i1}e^{x_{i1}\beta_0}y_{i1}d_{i2}\} \\
&= \frac{(x_{i1} - x_{i2})}{e^{v_0}}E\{x_{i2}e^{v_0 + x_{i2}\beta_0}y_{i2}d_{i1} - x_{i1}e^{v_0 + x_{i1}\beta_0}y_{i1}d_{i2}\} \\
&= \frac{(x_{i1} - x_{i2})}{e^{v_0}}E\{x_{i2}d_{i2}d_{i1} - x_{i1}d_{i1}d_{i2}\} \\
&= \frac{(x_{i1} - x_{i2})^2}{e^{v_0}}E\{d_{i2}d_{i1}\} \\
&= \frac{(x_{i1} - x_{i2})^2}{e^{v_0}}\Pr(d_{i1} = 1)\Pr(d_{i2} = 1).
\end{aligned}
$$

Calculation of the Information Matrix:

$$
\begin{aligned}
I &= \operatorname{var}\dot{Q}(\beta_0) = E[(x_{i1} - x_{i2})^2\{e^{x_{i2}\beta}y_{i2}d_{i1} - e^{x_{i1}\beta}y_{i1}d_{i2}\}^2] \\
&= (x_{i1} - x_{i2})^2 E[\{e^{x_{i2}\beta}y_{i2}d_{i1} - e^{x_{i1}\beta}y_{i1}d_{i2}\}^2] \\
&= (x_{i1} - x_{i2})^2 E[(e^{x_{i2}\beta}y_{i2}d_{i1})^2 - 2(e^{x_{i2}\beta}y_{i2}d_{i1})(e^{x_{i1}\beta}y_{i1}d_{i2}) + (e^{x_{i1}\beta}y_{i1}d_{i2})^2]
\end{aligned}
$$

where

$$
\begin{aligned}
E[(e^{x_{i2}\beta}y_{i2}d_{i1})^2] &= E[(e^{x_{i2}\beta}y_{i2})^2]E[(d_{i1})^2] \\
&= \frac{1}{e^{2v_0}}E[(e^{x_{i2}\beta + v_0}y_{i2})^2]E[(d_{i1})]
\end{aligned}
$$

$z_{is} = e^{x_{is}\beta + v_0}y_{is}$ has a Unit Exponential distribution, censored at $z_{is\,max} = e^{x_{is}\beta + v_0}c_{is}$:

$$Ez_{is}^2 = 2\{1 - e^{-z_{is\,\max}} - z_{is\,\max}e^{-z_{is\,\max}}\}$$

And therefore

$$\begin{aligned}
E[(e^{x_{i2}\beta}y_{i2}d_{i1})^2] &= \frac{1}{e^{2v_0}}E[(z_{i2})^2]E[(d_{i1})] \\
&= \frac{2}{e^{2v_0}}\{1 - e^{-z_{i2\,\max}} - z_{i2\,\max}e^{-z_{i2\,\max}}\}(1 - e^{-z_{i1\,\max}}).
\end{aligned}$$

Similar,

$$\begin{aligned}
E(e^{x_{i2}\beta}y_{i2}d_{i1})(e^{x_{i1}\beta}y_{i1}d_{i2}) &= E(e^{x_{i1}\beta}y_{i1}d_{i1})E(e^{x_{i2}\beta}y_{i2}d_{i2}) \\
&= \frac{1}{e^{2v_0}}E(z_{i1}d_{i1})E(z_{i2}d_{i2})
\end{aligned}$$

$$\begin{aligned}
E(z_{is}d_{is}) &= E(z_{is}|d_{is} = 1)\Pr(d_{is} = 1) \\
&= \int_0^{z_{is\,\max}} \frac{ze^{-z}}{1 - e^{-z_{is\,\max}}}dz \cdot (1 - e^{-z_{is\,\max}}) \\
&= \int_0^{z_{is\,\max}} ze^{-z}dz \\
&= \{1 - e^{-z_{i2\,\max}} - z_{i2\,\max}e^{-z_{i2\,\max}}\}.
\end{aligned}$$

Now we can calculate the Information Matrix:

$$\begin{aligned}
I &= \operatorname{var} Q_\beta(\beta_0) \\
&= \left(\frac{x_{i1} - x_{i2}}{e^{v_0}}\right)^2 E[(z_{i2}d_{i1})^2 - 2(z_{i2}d_{i1})(z_{i1}d_{i2}) + (z_{i1}d_{i2})^2] \\
&= 2\left(\frac{x_{i1} - x_{i2}}{e^{v_0}}\right)^2 \left\{1 + \frac{z_{i1\,\max}z_{i2\,\max}e^{-(z_{i1\,\max} + z_{i2\,\max})}}{(1 - e^{-z_{i1\,\max}})(1 - e^{-z_{i2\,\max}})}\right\}.
\end{aligned}$$

9 Literature

Blanchard, O.J., and L. F. Katz (1997): “What we Know and do Not Know about the Natural Rate of Unemployment,” Journal of Economic Perspectives, 11, 51-57.

Blanchard, O. J., and H. L. Summers (1988): “Why is Unemployment so High in Europe? Beyond the Natural Rate Hypothesis,” American Economic Review, 78, 182-187.

Blanchard, O. J., and J. Wolfers (1999): “The Role of Shocks and Institutions in the Rise of European Unemployment: The Aggregate Evidence,” Working paper, MIT.

Chamberlain, G. (1985): “Heterogeneity, Omitted Variable Bias, and Duration Dependence,” in Longitudinal Analysis of Labor Market Data, ed. by J. J. Heckman and B. Singer. Cambridge: Cambridge University Press.

Cox, D. R., and N. Reid (1987): “Parameter Orthogonality and Approximate Conditional Inference (with Discussion),” Journal of the Royal Statistical Society, Series B, 49, 1-39.

Cox, D. R., and N. Reid (1993): \A Note on the Calculation of Adjusted Pro¯le Likelihood," Journal of the Royal Statistical Society, Series B, 55, 467-471.

Dolton, P., and D. O'Neill (1996): \Unemployment Duration and the Restart E®ect: Some Experimental Evidence," The Economic Journal, 106, 387-400.

Heckman, J. J. (1991): \Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity," American Economic Review, 81, 75-79.

Heckman, J. J., and G. J. Borjas (1980): \Does Unemployment Cause Future Unemployment? De¯nitions, Questions and Answers for a Continuous Time Model of Heterogeneity and State Dependence," Economica, 47, 247-283.

Heckman, J. J., and B. Singer (1982): \The Identi¯cation Problem in Econometric Models for Duration Data," in Advances in Econometrics, ed. by W. Hildenbrand. New York: Cambridge University Press.

Heckman, J. J., and B. Singer (1984): \A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," Econometrica, 52, 271-320.

Honor¶e, B. E. (1990): \Simple Estimation of a Duration Model with Unobserved Heterogeneity," Econometrica, 58, 453-473.

Honor¶e, B. E. (1993): \Identi¯cation Results for Duration Models with Multiple Spells," Review of Economic Studies, 60, 241-246.

Honor¶e, B. E. (1998): \A Note of the Rate of Convergence of Estimators of Mixtures of Weibulls," Working paper, Princeton University.

Honor¶e, B. E., and J. L. Powell (1998): \Pairwise Di®erence Estimators for Non-Linear Models," Working paper, Princeton University.

Lancaster, T. (1976): \Redundancy, Unemployment and Manpower Policy: a Comment," Economic Journal, 86, 335-338.

Lancaster, T. (1979): \Econometric Methods for the Duration of Unemployment," Econometrica, 47, 939-956.

Lancaster, T. (1985): \Generalized 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: Cambridge University Press.

Lancaster, T. (1997): \Orthogonal Parameters and Panel Data," Working paper, Brown University.

Lancaster, T. (1999a): \The Incidental Parameters since 1948," Journal of Econometrics, forthcoming.

Lancaster, T. (1999b): \Some Econometrics of Scarring," Brown manuscript, Brown University.

Mundlak, Y. (1961): \Empirical Production Function Free of Management Bias," Journal of Farm Economics, 43, 44-56.

Newey, W. K., and D. McFadden (1994): \Large Sample Estimation and Hypothesis Testing," in Handbook of Econometrics, Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.

Neyman, J., and E. L. Scott (1948): \Consistent Estimation from Partially Consistent Observations," Econometrica, 16, 1-32.

Powell, J.L. (1994): \Estimation of Semiparametric Models," in Handbook of Econometrics, Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.

Ridder, G., and I. Tunali (forthcoming): \Strati¯ed Partial Likelihood Estimation," Journal of Econometrics.

Van den Berg, G. J. (forthcoming): \Duration Models: Speci¯cation, Identi¯cation, and Multiple Duration," in Handbook of Econometrics, Vol. 5, forthcoming. Amsterdam: North-Holland.

Woutersen, T. M. (1999): \Consistent Inference and Orthogonality" Brown manuscript, Brown University.

Yamaguchi, K. (1986): \Alternative approaches to unobserved heterogeneity in the analysis of repeatable events" in Sociological Methodology, Vol. 16, ed. by N. B. Tuma. San Francisco: Jossey-Bass.