Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 1014–1043
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Discrete time duration models with group-level heterogeneity

Anders Frederiksen[a,b], Bo E. Honoré[c,*], Luojia Hu[d]

[a]Hoover Institution, Stanford University, USA
[b]IZA and The Aarhus School of Business, Prismet, Silkeborgvej 2 DK-8000, Aarhus C, Denmark
[c]Department of Economics, Princeton University, Princeton, NJ 08544-1021, USA
[d]Department of Economics and Institute for Policy Research, Northwestern University, Evanston, IL 60208-2600, USA

Available online 7 February 2007

Abstract

Dynamic discrete choice panel data models have received a great deal of attention. In those models, the dynamics is usually handled by including the lagged outcome as an explanatory variable. In this paper we consider an alternative model in which the dynamics is handled by using the duration in the current state as a covariate. We propose estimators that allow for group-specific effect in parametric and semiparametric versions of the model. The proposed method is illustrated by an empirical analysis of job durations allowing for firm-level effects.

Keywords: Panel data; Discrete choice; Duration models

1. Introduction

Dynamic discrete choice panel data models have received a great deal of attention in statistics and econometrics. In those models, the dynamics is usually handled by including the lagged outcome as an explanatory variable. See for example Cox (1958), Heckman (1981a–c), Chamberlain (1985) and Honoré and Kyriazidou (2000). In the spirit of classical duration models where the dynamics is captured through dependence of the hazard on time (see Kalbfleisch and Prentice, 1980; Lancaster, 1990), this paper

*Corresponding author.
E-mail addresses: afr@asb.dk (A. Frederiksen), honore@Princeton.edu (BoE. Honoré), luojiahu@Northwestern.edu (L. Hu).

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.12.003

considers an alternative dynamic discrete choice model in which the dynamics is handled by using the duration in the current state as a covariate. Such a model can be interpreted as a discrete time duration model. The main contribution of the paper is to propose estimators that allow for group-specific effect in parametric and semiparametric versions of such a model. Duration models with group-specific effects have a long history, see for example Clayton and Cuzick (1985), Holt and Prentice (1974), Sastry (1997), Ridder and Tunali (1999) and Hougaard (2000). Most of these papers consider a parametric approach in which one assumes a distribution for the group-specific effects. A notable exception is the "fixed effects" approach in Ridder and Tunali (1999) who consider a conditioning approach similar to one that leads to Cox's partial likelihood estimator (Cox, 1972, 1975). Their approach works when durations are continuous, but breaks down if one has interval observations from the same model.

The starting point for this paper is to explicitly model the exit probabilities in a discrete time duration model. This is different from deriving the exit probabilities from an underlying continuous time model. The advantage of this is that we are able to incorporate group-specific effects in the spirit of Ridder and Tunali (1999) in a discrete time duration model.

Heckman (1981a–c), Honoré and Kyriazidou (2000) and others studied a dynamic panel data model of the type

$$y_{it} = 1\{x_{it}'\beta + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it} \geqslant 0\} \tag{1}$$

where the explanatory variables, x_{it}, are strictly exogenous under various assumptions of the distribution of ε_{it}. This model is empirically relevant in many situations. Specifically, the term α_i can be thought of as capturing unobserved heterogeneity; some individuals are consistently more likely to experience the event than others. The term, $\gamma y_{i,t-1}$, captures state dependence; the probability that an individual experiences the event this period depends on whether the event happened last period. See e.g., Heckman (1981c). While both unobserved heterogeneity and state dependence are important, (1) ignores a third source of persistence, namely duration dependence. In duration models, duration dependence refers to the phenomenon that the time since the last occurrence of the event might affect the probability that the event occurs now. Clearly the time since the last occurrence of the event is not strictly exogenous, and the approach in Honoré and Kyriazidou (2000) will not work if it is included in x_{it}.

In Section 2 below, we define the model and propose estimators under alternative assumptions. We also make a link to the estimation of monotone index models and continuous time duration models. Section 3 considers multiple-spell versions of the model. Here one has to distinguish between two cases. Sometimes it is reasonable to assume that the spells are drawn from the same distribution. One example of this would be time between purchases of identical products. In other situations, consecutive spells are clearly drawn from different distributions. For example, one worker can alternate between employment and unemployment spells. Section 4 applies the approach developed in Section 2 to analyze job durations using a unique Danish data set. This application confirms that it is important to control for group-specific effects. Section 5 concludes.

2. The model and estimator

The maintained assumption in this paper is that we observe a sample of individuals that is grouped in such a way that the individual–specific effect is the same within the group.[1] We will use i to denote a group and j to index individuals within a group. We will assume that the number of groups is large relative to the number of time–periods and the number of individuals within each group. The relevant asymptotic is therefore one that assumes that the number of groups increases.

In this section we focus on single spell models. Since some spells will be in progress at the start of the sampling process, the time at which a spell ends will not necessarily equal the duration of the spell. It is therefore necessary to define a number of variables related to the duration of the spell. For each individual, we use S_{ji1} to denote the duration of the spell at the beginning of the sample period, and we use T_{ji} to denote the sampling period in which the spell ends. This means that the duration of the spell for individual j in group i will be $\Upsilon_{ji} = S_{ji1} + T_{ji}$.

We formulate the model as a modification of the dynamic discrete choice model in (1) in which the lagged dependent variable has been replaced by the number of periods since the individual entered the state of interest. $y_{jit} = 1$ will be used to describe the event that individual j in group i leaves the state at calendar time t. Hence the model is

$$y_{jit} = 1\{x'_{jit}\beta + \delta_{S_{jit}} + \alpha_i + \varepsilon_{jit} \geqslant 0\}, \quad t = 1, \ldots, \bar{t}, \;\; j = 1, \ldots, J_i, \;\; i = 1, \ldots, n, \tag{2}$$

where S_{jit} denotes the duration of the spell at time t (i.e., $S_{jit} = S_{ji1} + t$). $\bar{t}$ is the end of the sampling period. We will use y_i and y_{ji} to denote $\{y_{jit} : t = 1, \ldots, \bar{t}, j = 1, \ldots, J_i\}$ and $\{y_{jit} : t = 1, \ldots, \bar{t}\}$, respectively. Similar notation will be used for the explanatory variables x. It is also not necessary that one observes data for an individual after the event has occurred. This is for example relevant if T_{ji} is the time at which some failure (such as death) occurs. We will therefore assume that we observe $\{x'_{jit} : t = 1, \ldots, T_{ji}, j = 1, \ldots, J_i, i = 1, \ldots, n\}$, and we need only assume that (2) applies for $t = 1, \ldots, T_{ji}$.

It is clear that a scale normalization is needed for estimation of (β, δ), and that a location normalization is needed on the duration dependence parameter δ's. In some applications, it might also be natural to restrict the "duration dependence" by, for example, assuming that δ_t is a linear function of t.

The model in (2) is relevant when one worries about an unobserved heterogeneity component which is the same for all individuals in a group. This situation will for example emerge if one has a sample of workers where some of them work in the same firm and where one wants to control for firm-specific effects. A second example is the case where one observes individual members of a household and wants to control for household-specific effects. In the spirit of "fixed effects" panel data models, we will not restrict the distribution of the group-specific effect, α_i, and we do not assume that it is independent of the strictly exogenous variables x_{jit}. Whether a random effects approach is more desirable is application specific. If it is, then parametric versions of the model can be estimated using textbook classical or Bayesian methods. One situation in which a random effects approach is typically undesirable, is when the first observation in the sample does not correspond to the first period that the individual is in the state. This is due to the usual left

[1]This is sometimes referred to as parallel data (see e.g., Hougaard, 2000) although it is not necessary that observations in the same group enter the state at the same point in time.

censoring/initial conditions problem that occurs when some spells are in progress at the start of the sampling process.

In what follows, we will assume that the number of observations in a group, J_i, is the same across groups. This can be easily relaxed provided that J_i is exogenous (formally, the assumptions below have to hold conditional on J_i). In the spirit of linear panel data models, the proposed estimation technique will be based on the observations for which $J_i \geqslant 2$.

We assume that we have a random sample of groups indexed by i.

Assumption 1. All random variables corresponding to different i are independent of each other and identically distributed.

We consider three versions of the model. The three differ in the assumptions that are made on the distribution of ε_{jit}. To state the assumptions formally and in some generality, we define z_i to be all the predetermined characteristics of the group at the beginning of the sample. These will include α_i, $\{x_{ki1}\}_{k=1}^{J}$, $\{S_{ki1}\}_{k=1}^{J}$ as well as characteristics of the group that do not enter the model directly.

Assumption 2a. For each i and t, the ε_{jit}'s are all logistically distributed conditional on $\{\alpha_i, \{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leqslant t}, \{\varepsilon_{kis}\}_{s\leqslant t+\tau, k\neq j}, \{x_{kis}\}_{s\leqslant t+\tau, k\neq j}, \{S_{ki1}\}_{k=1}^{J}\}$ for some known $\tau \geqslant 0$.

This assumption corresponds to the logit assumption used in Rasch (1960), Cox (1958), Andersen (1970), Chamberlain (1985), Honoré and Kyriazidou (2000), Thomas (2002) and others. For a given individual, Assumption 2a does not limit the feedback from the ε's to future values of x. The setup therefore allows x to be predetermined. As a result, there is no need to treat $\delta_{S_{jit}}$ in (2) differently from the other explanatory variables. However, the notation in (2) makes it easier to compare the approach here to literature, and the duration dependence may be of special interest.

However, when $\tau > 0$, it is assumed that a "feedback" from one individual's ε to the other group member's x's and ε's is nonexistent for τ periods. τ is therefore application specific.

The next assumption generalizes Assumption 2a by allowing ε_{jit} to have an unknown, but common, distribution. This is in the spirit of the way in which Manski (1987) generalized Rasch's logit model with individual-specific effects.

Assumption 2b. For some known τ ($\tau \geqslant 0$), and conditionally on z_i, $\{\varepsilon_{jit}\}_{j=1}^{J}$ are independent of each other and of $\{\{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leqslant t}, \{\varepsilon_{kis}\}_{s\leqslant t+\tau, k\neq j}, \{x_{kis}\}_{s\leqslant t+\tau, k\neq j}\}$ for $t = 1, \ldots, T$, and the conditional distributions of $\{\varepsilon_{jit}\}_{j=1,t=1}^{J,T}$ are identical.

Note that under Assumption 2b, the distributions of ε_{jit} are allowed to vary across i.

Assumptions 2a and 2b fit naturally with the assumptions that are made in the discrete choice literature. Moreover, Assumption 2b can be interpreted as the result of having interval observations from a standard continuous time proportional hazard model with piecewise constant explanatory variables. See Section 2.5.

In Assumption 2c below we allow the distribution of ε_{jit} to depend on S_{jit}.

Assumption 2c. For some known τ ($\tau \geqslant 0$), and conditionally on z_i, $\{\varepsilon_{jit}\}_{j=1}^{J}$ are independent of each other and of $\{\{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leqslant t}, \{\varepsilon_{kis}\}_{s\leqslant t+\tau, k\neq j}, \{x_{kis}\}_{s\leqslant t+\tau, k\neq j}\}$. Moreover, the distributions of ε_{jit} and $\varepsilon_{\ell is}$ are identical if s and t correspond to the same duration time.

It is clear that Assumption 2c is weaker than Assumption 2b. This will, in itself, make it interesting to consider Assumption 2c. However, the main motivation for Assumption 2c is that it allows us to make a connection between the models considered here and the monotone index model, and hence implicitly with proportional hazard models and with accelerated failure time models. Specifically, consider the model

$$G(T^*_{ji}) = x'_{ji}\beta + \alpha_i + \varepsilon_{ji}, \tag{3}$$

where G is continuous and strictly increasing and a discretized version of T^*_{ji} is observed. Eq. (3) implies that

$$\mathrm{P}(T^*_{ji} > t+1 | x_{ji}, T^*_{ji} > t) = \frac{1 - F(G(t+1) - x'_{ji}\beta - \alpha_i)}{1 - F(G(t) - x'_{ji}\beta - \alpha_i)},$$

where F is the CDF for ε_{ji}. When $1 - F(\cdot)$ is log-concave (which is implied by the density of ε_i being log-concave; see Heckman and Honoré , 1990), the right-hand side is an increasing function of $x'_{ji}\beta + \alpha_i$. This means that one can write the event $T^*_{ji} > t+1 | x_{ji}, T^*_{ji} > t$ in the form $1\{x'_{ji}\beta + \alpha_i > \eta_{jit}\}$ for some random variable η_{jit} which is independent of x_{ji} and has CDF $(1 - F(G(t+1) - \cdot))/(1 - F(G(t) - \cdot))$. This has the same structure as (2) with time–invariant explanatory variables combined with Assumption 2c. In other words, a monotone index model with discretized observations of the dependent variable and log-concave errors, is a special case of the model considered here. See also Frederiksen et al. (2006).

For now assume that $J = 2$. The following lemma is crucial for the results in this paper.

Lemma 1. *Let t_1 and t_2 be arbitrary with $|t_1 - t_2| \leqslant \tau$. Consider the two events $A = \{T_{1i} = t_1, T_{2i} > t_2\}$ and $B = \{T_{1i} > t_1, T_{2i} = t_2\}$. Under Assumption* 2a

$$\mathrm{P}(A | A \cup B, x_{1it_1}, x_{2it_2}, z_i) = \frac{\exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))}{1 + \exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))},$$

under Assumption 2b

$$\mathrm{P}(A | A \cup B, x_{1it_1}, x_{2it_2}, z_i) \begin{cases} > \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) > 0, \\ = \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) = 0, \\ < \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) < 0 \end{cases}$$

and under Assumption 2c *and if $t_1 + S_{1i1} = t_2 + S_{2i1}$,*

$$\mathrm{P}(A | A \cup B, x_{1it}, x_{2it}, z_i) \begin{cases} > \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta > 0, \\ = \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta = 0, \\ < \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta < 0. \end{cases}$$

Lemma 1 suggests estimators of β and δ. Under Assumption 2a, one can estimate β and $\{\delta_t\}$ by maximizing

$$\sum_{i=1}^{n} \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1\{|t_1 - t_2| \leqslant \tau\}(1\{T_{1i} = t_1, T_{2i} > t_2\} + 1\{T_{1i} > t_1, T_{2i} = t_2\})$$
$$\times \log\left(\frac{\exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))^{1\{T_{1i}=t_1, T_{2i}>t_2\}}}{1 + \exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))}\right). \tag{4}$$

This estimator is a standard extremum estimator and consistency and asymptotic normality (as n increases to infinity) are easily established (using for example the arguments in Amemiya, 1985). Specifically,

$$\sqrt{n}(\widehat{\theta} - \theta_0) \widetilde{\longrightarrow} \mathrm{N}(0, A^{-1}BA^{-1}),$$

where $\widehat{\theta} = (\widehat{\beta}', \widehat{\delta}')'$, $\theta_0 = (\beta_0', \delta_0')'$,

$$A = \mathrm{E}\left[\left. \frac{\partial^2 q_i(\theta)}{\partial\theta\partial\theta'} \right|_{\theta_0} \right]$$

and

$$B = \mathrm{E}\left[\left. \frac{\partial q_i(\theta)}{\partial\theta} \right|_{\theta_0} \left. \frac{\partial q_i(\theta)}{\partial\theta} \right|_{\theta_0}' \right],$$

where

$$\begin{aligned} q_i(\theta) = & \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1\{|t_1 - t_2| \leqslant \tau\}(1\{T_{1i} = t_1, T_{2i} > t_2\} + 1\{T_{1i} > t_1, T_{2i} = t_2\}) \\ & \times \log\left(\frac{\exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))^{1\{T_{1i}=t_1, T_{2i}>t_2\}}}{1 + \exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))} \right). \end{aligned}$$

Similarly, under Assumption 2b, one can estimate β and $\{\delta_t\}$ (up to scale) by a maximum score estimator in the spirit of Manski (1975, 1987). Specifically this estimator would maximize

$$\begin{aligned} & \sum_{i=1}^{n} \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1\{|t_1 - t_2| \leqslant \tau\} \cdot 1\{T_{1i} = t_1, T_{2i} > t_2\} \\ & \quad \times 1\{(x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) > 0\} \\ & \qquad + 1\{|t_1 - t_2| \leqslant \tau\} \cdot 1\{T_{1i} > t_1, T_{2i} = t_2\} \\ & \qquad \times 1\{(x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) < 0\} \end{aligned} \tag{5}$$

subject to a scale normalization. Following the arguments in Manski (1975, 1987), this estimator is consistent under random sampling subject to support conditions on the distribution of the explanatory variables. Similar to Horowitz (1992), a smoothed maximum score estimator defined by maximization of a smoothed version of (5) will be asymptotically normal (although its rate of convergence will be slower than the usual $\sqrt{n}$).

Finally, under Assumption 2c, one can estimate β (up to scale) by maximizing

$$\begin{aligned} & \sum_{i=1}^{n} \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1\{t_1 + S_{1i1} = t_2 + S_{2i1}\} \cdot 1\{|t_1 - t_2| \leqslant \tau\} \\ & \quad \times (1\{T_{1i} = t_1, T_{2i} > t_2\} \cdot 1\{(x_{1it_1} - x_{2it_2})'\beta > 0\} \\ & \quad + 1\{T_{1i} > t_1, T_{2i} = t_2\} \cdot 1\{(x_{1it_1} - x_{2it_2})'\beta < 0\}) \end{aligned} \tag{6}$$

subject to a scale normalization. In this case, the δ's are not identified. This is because Assumption 2c places no restriction on the location of ε.

In the discussion leading up to Lemma 1 and Eqs. (4)–(6), we implicitly assume that the calendar time for the first observation is the same for all individuals. If this is not the case, then the feedback in Assumptions 2a–2c should refer to the calendar time rather than the duration time. As a result, the statement $|t_1 - t_2| \leqslant \tau$ should be replaced by a statement that the calendar times are within τ, and indicator functions $1\{|t_1 - t_2| \leqslant \tau\}$ in Eqs. (4)–(6) should be replaced by indicator functions for the difference in the calendar times being less than or equal to τ.

2.1. Group-specific δ or x

Note that the δ-terms drop out in the case where $t_1 + S_{1i1} = t_2 + S_{2i1}$ in Lemma 1. This allows us to construct an estimator for the case where δ_t is also indexed by i by only including terms for which $t_1 + S_{1i1} = t_2 + S_{2i1}$ in (4)–(6). This is similar in spirit to the continuous time panel duration model considered by Ridder and Tunali (1999) (see below). It is also somewhat similar to the approach in Chamberlain (1985) and D'Adddio and Honoré (2006). Those papers consider models with second order state dependence where the first order is allowed to be individual specific.

It is also worth noting that if τ in Assumptions 2a–2c is positive, then the approach taken here allows us to estimate a model in which all the explanatory variables are group specific, $x_{1it} = x_{2it}$ for all t. Conversely, if $\tau = 0$ then all group-specific terms will cancel in (4)–(6). This implies that we can allow for group-specific, time-varying shocks.

2.2. Censoring

Covariate-dependent censoring is not a problem provided that it is independent of the ε's. Specifically, assume that we observe $\{y_{jit}, x_{jit}\}$ only up to (and including) some random period C_{ji}. In other words, C_{ji} is the censoring time for T_{ji} (measured in "sample" time) and with the convention that it is observed whether the event $T_{ji} = C_{ji}$ occurs.

The argument above then applies if Assumptions 2a–2c are modified to:

Assumption 2a′. For each i and t, the ε_{jit}'s are all logistically distributed conditional on $\{\alpha_i, \{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leqslant t}, \{\varepsilon_{kis}\}_{s\leqslant t+\tau,k\neq j}, \{x_{kis}\}_{s\leqslant t+\tau,k\neq j}, \{S_{ki1}\}_{k=1}^{J}, \{C_{ki}\}_{k=1}^{J}\}$ for some known τ.

Assumption 2b′. For some known τ ($\tau \geqslant 0$), and conditionally on z_i, $\{\varepsilon_{jit}\}_{j=1}^{J}$ are independent of each other and of $\{\{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leqslant t}, \{\varepsilon_{kis}\}_{s\leqslant t+\tau,k\neq j}, \{x_{kis}\}_{s\leqslant t+\tau,k\neq j}, \{C_{ki}\}_{k=1}^{J}\}$ for $t = 1, \ldots, T$, and the conditional distributions of $\{\varepsilon_{jit}\}_{j=1,t=1}^{J,T}$ are identical.

Assumption 2c′. For some known τ ($\tau \geqslant 0$), and conditionally on z_i, $\{\varepsilon_{jit}\}_{j=1}^{J}$ are independent of each other and of $\{\{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leqslant t}, \{\varepsilon_{kis}\}_{s\leqslant t+\tau,k\neq j}, \{x_{kis}\}_{s\leqslant t+\tau,k\neq j}, \{C_{ki}\}_{k=1}^{J}\}$. Moreover, the distributions of ε_{jit} and $\varepsilon_{\ell is}$ are identical if s and t correspond to the same duration time.

Hence under Assumption 2a′, one can estimate β and $\{\delta_t\}$ by maximizing

$$\sum_{i=1}^{n}\sum_{t_1=1}^{\bar{t}}\sum_{t_2=1}^{\bar{t}} 1\{|t_1 - t_2| \leqslant \tau, t_1 < C_{1i}, t_2 < C_{2i}\}$$
$$\times (1\{T_{1i} = t_1, T_{2i} > t_2\} + 1\{T_{1i} > t_1, T_{2i} = t_2\})$$
$$\times \log\left(\frac{\exp((x_{1it_1} - x_{2it_2})\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))^{1\{T_{1i}=t_1, T_{2i}>t_2\}}}{1 + \exp((x_{1it_1} - x_{2it_2})\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))}\right).$$

Similarly, under Assumption 2b′, one can estimate β and $\{\delta_t\}$ (up to scale) by maximizing

$$\sum_{i=1}^{n}\sum_{t_1=1}^{\bar{t}}\sum_{t_2=1}^{\bar{t}} 1\{|t_1-t_2|\leqslant\tau, t_1<C_{1i}, t_2<C_{2i}\}$$
$$\times 1\{T_{1i}=t_1, T_{2i}>t_2\}\cdot 1\{(x_{1it_1}-x_{2it_2})'\beta+(\delta_{t_1+S_{1i1}}-\delta_{t_2+S_{2i1}})>0\}$$
$$+\,1\{|t_1-t_2|\leqslant\tau, t_1<C_{1i}, t_2<C_{2i}\}$$
$$\times 1\{T_{1i}>t_1, T_{2i}=t_2\}\cdot 1\{(x_{1it_1}-x_{2it_2})'\beta+(\delta_{t_1+S_{1i1}}-\delta_{t_2+S_{2i1}})<0\}$$

subject to a scale normalization.

Finally, under Assumption 2c′, one can estimate β (up to scale) by maximizing

$$\sum_{i=1}^{n}\sum_{t_1=1}^{\bar{t}}\sum_{t_2=1}^{\bar{t}} 1\{|t_1-t_2|\leqslant\tau, t_1+S_{1i1}=t_2+S_{2i1}, t_1<C_{1i}, t_2<C_{2i}\}$$
$$\cdot(1\{T_{1i}=t_1, T_{2i}>t_2\}\times 1\{(x_{1it_1}-x_{2it_2})'\beta>0\}+1\{T_{1i}>t_1, T_{2i}=t_2\}\times 1\{(x_{1it_1}-x_{2it_2})'\beta<0\}).$$

2.3. More than two observations per unit

A similar approach can be used when there are more than two observations for each group.

To illustrate this, suppose that a group has three observations and define $A=\{T_{1i}=t_1, T_{2i}>t_2, T_{3i}>t_3\}$, $B=\{T_{1i}>t_1, T_{2i}=t_2, T_{3i}>t_3\}$ and $C=\{T_{1i}>t_1, T_{2i}>t_2, T_{3i}=t_3\}$. Under the logit Assumption 2a, we then have

$$\mathrm{P}(A|A\cup B\cup C, x_{1it_1}, x_{2it_2}, x_{3it_3})$$
$$=\frac{\exp(x'_{1it_1}\beta+\delta_{t_1+S_{1i1}})}{\exp(x'_{1it_1}\beta+\delta_{t_1+S_{1i1}})+\exp(x'_{2it_2}\beta+\delta_{t_2+S_{2i1}})+\exp(x'_{3it_3}\beta+\delta_{t_3+S_{3i1}})}.$$

For the semiparametric case in Assumption 2b, we get

$$\mathrm{P}(A|A\cup B\cup C, x_{1it_1}, x_{2it_2}, x_{3it_3})$$
$$>\max\{\mathrm{P}(B|A\cup B\cup C, x_{1it_1}, x_{2it_2}, x_{3it_3}), \mathrm{P}(C|A\cup B\cup C, x_{1it_1}, x_{2it_2}, x_{3it_3})\}$$

if and only if

$$x'_{1it_1}\beta+\delta_{t_1+S_{1i1}}>\max\{x'_{2it_2}\beta+\delta_{t_2+S_{2i1}}, x'_{3it_3}\beta+\delta_{t_3+S_{3i1}}\}.$$

This has the same structure as the multinomial qualitative response model of Manski (1975), and the insights there can be used to construct a maximum score estimator.

Under Assumption 2c, we can use the case where $t_1+S_{1i1}=t_2+S_{2i1}=t_3+S_{3i1}$ (so they all refer to the same duration) and we have

$$\mathrm{P}(A|A\cup B\cup C, x_{1it_1}, x_{2it_2}, x_{3it_3})$$
$$>\max\{\mathrm{P}(B|A\cup B\cup C, x_{1it_1}, x_{2it_2}, x_{3it_3}), \mathrm{P}(C|A\cup B\cup C, x_{1it_1}, x_{2it_2}, x_{3it_3})\}$$

if and only if

$$x'_{1it_1}\beta>\max\{x'_{2it_2}\beta, x'_{3it_3}\beta\}.$$

We could also define $A = \{T_{1i} = t_1, T_{2i} = t_2, T_{3i} > t_3\}$, $B = \{T_{1i} = t_1, T_{2i} > t_2, T_{3i} = t_3\}$ and $C = \{T_{1i} > t_1, T_{2i} = t_2, T_{3i} = t_3\}$. Under the logit Assumption 2a, we then have

$$\mathrm{P}(A|A \cup B \cup C, x_{1it_1}, x_{2it_2}, x_{3it_3}) = \frac{c_1}{c_1 + c_2 + c_3},$$

where

$$\begin{aligned} c_1 &= \exp((x_{1it_1} + x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} + \delta_{t_2+S_{2i1}})), \\ c_2 &= \exp((x_{1it_1} + x_{3it_3})'\beta + (\delta_{t_1+S_{1i1}} + \delta_{t_3+S_{3i1}})), \\ c_3 &= \exp((x_{2it_2} + x_{3it_3})'\beta + (\delta_{t_2+S_{2i1}} + \delta_{t_3+S_{3i1}})). \end{aligned}$$

For the semiparametric case in Assumption 2b, we get

$$\begin{aligned} &\mathrm{P}(A|A \cup B \cup C, x_{1it_1}, x_{2it_2}, x_{3it_3}) \\ &\quad > \max\{\mathrm{P}(B|A \cup B \cup C, x_{1it_1}, x_{2it_2}, x_{3it_3}), \mathrm{P}(C|A \cup B \cup C, x_{1it_1}, x_{2it_2}, x_{3it_3})\} \end{aligned}$$

if and only if

$$\begin{aligned} &(x_{1it_1} + x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} + \delta_{t_2+S_{2i1}}) \\ &\quad > \max\{(x_{1it_1} + x_{3it_3})'\beta + (\delta_{t_1+S_{1i1}} + \delta_{t_3+S_{3i1}}), (x_{2it_2} + x_{3it_3})'\beta + (\delta_{t_2+S_{2i1}} + \delta_{t_3+S_{3i1}})\}. \end{aligned}$$

This can be used to construct a maximum score estimator in the spirit of Manski (1975).

Under Assumption 2c, we can use the case where $t_1 + S_{1i1} = t_2 + S_{2i1} = t_3 + S_{3i1}$ (so they all refer to the same duration) and we have

$$\begin{aligned} &\mathrm{P}(A|A \cup B \cup C, x_{1it_1}, x_{2it_2}, x_{3it_3}) \\ &\quad > \max\{\mathrm{P}(B|A \cup B \cup C, x_{1it_1}, x_{2it_2}, x_{3it_3}), \mathrm{P}(C|A \cup B \cup C, x_{1it_1}, x_{2it_2}, x_{3it_3})\} \end{aligned}$$

if and only if

$$(x_{1it_1} + x_{2it_2})'\beta > \max\{(x_{1it_1} + x_{3it_3})'\beta, (x_{2it_2} + x_{3it_3})'\beta\}.$$

We can derive similar expression for $J > 3$. Alternatively, one could consider all pairs of observations within a group.

2.4. Conditional likelihood

Most of the existing results for logit models with individual-specific effects have been based on a conditional likelihood approach. A sufficient statistic, S_i, for α_i in (2) is defined to be a function of the data such that the distribution of y_i conditional on (S_i, x_i, α_i), does not depend on α_i. If one has a sufficient statistic, which furthermore has the property that the distribution of y_i conditional on (S_i, x_i, α_i) depends on the parameters of interest, then those can be estimated by maximum likelihood using the conditional distribution of the data, given the sufficient statistic. Andersen (1970) proved that the resulting estimator is consistent and asymptotically normal under appropriate regularity conditions. Unfortunately, it does not appear that the method proposed here can be motivated as a conditional likelihood estimator.

For simplicity assume that x_i is strictly exogenous. Under Assumption 2a, the distribution of y_i given (x_i, α_i) is then

$$\left(\prod_{s=1}^{T_{1i}-1}\frac{1}{1+\exp(x'_{1is}\beta+\delta_{S_{1is}}+\alpha_i)}\right)\frac{\exp(x'_{1iT_{1i}}\beta+\delta_{S_{1iT_{1i}}}+\alpha_i)}{1+\exp(x'_{1iT_{1i}}\beta+\delta_{S_{1iT_{1i}}}+\alpha_i)}$$
$$\times\left(\prod_{s=1}^{T_{2i}-1}\frac{1}{1+\exp(x'_{2is}\beta+\delta_{S_{2is}}+\alpha_i)}\right)\frac{\exp(x'_{2iT_{2i}}\beta+\delta_{S_{2iT_{2i}}}+\alpha_i)}{1+\exp(x'_{2iT_{2i}}\beta+\delta_{S_{2iT_{2i}}}+\alpha_i)}$$
$$=\frac{\exp(2\alpha_i)\exp(x'_{1iT_{1i}}\beta+\delta_{S_{1iT_{1i}}}+x'_{2iT_{2i}}\beta+\delta_{S_{2iT_{2i}}})}{\prod_{s=1}^{T_{1i}}(1+\exp(x'_{1is}\beta+\delta_{S_{1is}}+\alpha_i))\prod_{s=1}^{T_{2i}}(1+\exp(x'_{2is}\beta+\delta_{S_{2is}}+\alpha_i))}.$$

It follows from that the sufficient statistic is (T_{1i}, T_{2i}). Hence, a conditional likelihood approach will not work.

2.5. Comparison to continuous case

The hazard for the proportional hazard model with time-varying covariates is

$$\lambda(t|\{x_{is}\}_{s\leqslant t})=\lambda(t)\exp(x'_{it}\beta)$$

(see Kalbfleisch and Prentice, 1980). Cox's (1972, 1975) estimator essentially conditions on the failure times and, for each failure time, on the risk set (the set of observations that have not yet experienced the event and are not yet censored). The contribution to the "likelihood" function for an observation, i, that experiences the event at duration–time t is then the probability that, of the observations at risk at duration–time t, the ith is the one to experience the event (given that one of them will). For the proportional hazard model, this probability has the same functional form as a multinomial logit. This insight was used in Ridder and Tunali (1999) in the case where the observations are grouped in the way discussed here. The resulting estimator is based on an objective function which has terms similar to the contributions in (4) from $t_1 + S_{1i1} = t_2 + S_{2i1}$.

Prentice and Gloeckler (1978) and Meyer (1990) considered estimation in a proportional hazard model with interval data and piecewise constant explanatory variables. In that case

$$P(y_{jit}=1|\{x_{jis}\}_{s\leqslant t},\alpha_i)=P(T^*_{ji}<t|T^*_{ji}>t-1,\{x_{jis}\}_{s\leqslant t},\alpha_i),$$

where T^*_{ji} denotes the underlying continuous duration.

If T^*_{ji} has hazard

$$\lambda(t|\{x_{jis}\}_{s\leqslant t},\alpha_i)=\lambda(t)\exp(x'_{jit}\beta+\alpha_i)$$

then

$$\begin{aligned}P(y_{jit}=1|\{x_{jis}\}_{s\leqslant t},\alpha_i)&=1-\exp\left(-\int_{t-1}^{t}\lambda(s)\exp(x'_{jis}\beta+\alpha_i)\,\mathrm{d}s\right)\\&=1-\exp\left(-\exp(x'_{jit}\beta+\alpha_i)\int_{t-1}^{t}\lambda(s)\,\mathrm{d}s\right)\\&=1-\exp(-\exp(x'_{jit}\beta+\delta_t+\alpha_i)),\end{aligned}$$

where

$$\delta_t=\log\left(\int_{t-1}^{t}\lambda(s)\,\mathrm{d}s\right).$$

In other words (after allowing for left censoring), one can write
$y_{jit} = 1\{x'_{jit}\beta + \delta_{S_{jit}} + \alpha_i + \varepsilon_{jit} \geqslant 0\}$,

where ε_{jit} is Type–1 extreme value distributed (i.e., has CDF $F(\eta) = \exp(-\exp(-\eta))$). In other words, the proportional hazard model with interval data fits our setup with Assumption 2b.

Finally, we note that it is possible to interpret the model that results from Assumption 2a, as the outcome of a proportional hazard model with i.i.d. piecewise shocks to the hazard. Specifically, assume that the hazard for T^*_{ji} is

$$\lambda(t|\{x_{jis}\}_{s\leqslant t}, \alpha_i, \{v_{jis}\}_{s\leqslant t}) = \lambda(t)\exp(x'_{jit}\beta + \alpha_i - v_{jit}),$$

where v_{jit} is constant over each time interval, and is i.i.d. extreme value distributed. Then

$$\mathrm{P}\Big(y_{jit} = 1|\{x_{jis}\}_{s\leqslant t}, \alpha_i, \{v_{jis}\}_{s\leqslant t}\Big) = 1 - \exp(-\exp(x'_{jit}\beta + \delta_t + \alpha_i - v_{jit}))$$

so one can write

$$y_{jit} = 1\{x'_{jit}\beta + \delta_{S_{jit}} + \alpha_i + \varepsilon_{jit} - v_{jit} \geqslant 0\}. \tag{7}$$

Since the difference in two extreme value distributed random variables is logistic, it follows that (7) is the model that results from Assumption 2a.

3. Multiple-spell versions of the model

The previous section considered single spell models. This is reasonable in situations where the event is one that can happen only once. On the other hand, there are many situations in which the event can reoccur. For example, one might want to model the duration between purchases of a particular good. In that case it would be reasonable to assume that the process starts over at the end of each spell. There are also cases that fall in between these extremes. One example of that could be the timing of births. In this case, the spell between the first and second child starts at the point when the first child is born. This is similar to the case of an individual purchasing a good. On the other hand, it may not be reasonable to specify the same model for, for example, the duration between the birth of the first and second child as one would for the duration between the birth of the third and fourth child. A two–state discrete time duration model is also an "intermediate case."

In this section, we discuss how the ideas in the previous section generalize to multiple-spell models. The derivations are given in the Appendix (see Section A.2).

3.1. Models with two spells

To fix ideas, we augment the setup in the previous section by assuming that a new spell of a potentially different type starts when the first spell ends. To accommodate this in the notation, we use superscript 1 for the first duration and superscript 2 for the second duration.

The model then is

$$y^1_{jit} = 1\{x'_{jit}\beta^1 + \delta^1_{S^1_{jit}} + \alpha^1_i + \varepsilon_{jit} \geqslant 0\}, \quad t = 1, \ldots, \bar{t},\ j = 1, \ldots, J,\ i = 1, \ldots, n,$$

$$jit^2 = 1\{x'_{jit}\beta^2 + \delta^2_{S^2_{jit}} + \alpha^2_i + \varepsilon_{jit} \geqslant 0\}, \quad t = T^1_{ji} + 1, \ldots, \bar{t},\ j = 1, \ldots, J\ \ i = 1, \ldots, n.$$

This notation allows the two spells to be fundamentally different (e.g., a spell of employment followed by a spell of unemployment) and the case where they are of the same type is the special case in which all parameters in the two equations are the same.

For notational simplicity, we consider only the case where $J = 2$.

3.1.1. Comparing first spells

One can use the first spells of individuals i_1 and i_2 to construct conditional statements like the ones in the previous section to estimate β^1 and δ^1.

3.1.2. Comparing first spells to second spells (assuming $\alpha_i^1 = \alpha_i^2 = \alpha_i$)

In this subsection we illustrate that it is possible to construct probability statements that are informative about the parameters of interest by comparing the first spell for one individual to the second spell for a different individual. This requires that the group-specific effect does not depend on the spell number. Whether or not this is reasonable depends on the empirical application that one has in mind.[2]

Let t_1^1, t_1^2 and t_2^1 be arbitrary with $t_1^1 < t_1^2$ and $|t_1^2 - t_2^1| \leqslant \tau$. Consider the two events $A = \{T_{1i}^1 = t_1^1, T_{1i}^2 = t_1^2, T_{2i}^1 > t_2^1\}$ and $B = \{T_{1i}^1 = t_1^1, T_{1i}^2 > t_1^2, T_{2i}^1 = t_2^1\}$.

In the Appendix we show that under Assumption 2a

$$\mathrm{P}(A|A \cup B, x_{1it_1^2}, x_{2it_2^1}, z_i) = \frac{\exp(x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta^2_{t_1^2 - t_1^1} - \delta^1_{t_2^1 + S_{2i1}})}{1 + \exp(x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta^2_{t_1^2 - t_1^1} - \delta^1_{t_2^1 + S_{2i1}})} \tag{8}$$

and under Assumption 2b

$$\mathrm{P}\Big(A|A \cup B, x_{1it_1^2}, x_{2it_2^1}, z_i\Big) \begin{cases} > \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta^2_{t_1^2 - t_1^1} - \delta^1_{t_2^1 + S_{2i1}} > 0, \\ = \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta^2_{t_1^2 - t_1^1} - \delta^1_{t_2^1 + S_{2i1}} = 0, \\ < \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta^2_{t_1^2 - t_1^1} - \delta^1_{t_2^1 + S_{2i1}} < 0. \end{cases} \tag{9}$$

Finally, under Assumption 2c, and if $t_1^2 - t_1^1 = t_2^1 + S_{2i1}$,

$$\mathrm{P}\Big(A|A \cup B, x_{1it_1^2}, x_{2it_2^1}, z_i\Big) \begin{cases} > \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 > 0, \\ = \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 = 0, \\ < \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 < 0. \end{cases} \tag{10}$$

Since (8)–(10) do not depend on t_1^1, the same statements are true if we redefine A and B as $A = \{T_{1i}^2 = t_1^2, T_{2i}^1 > t_2^1\}$ and $B = \{T_{1i}^2 > t_1^2, T_{2i}^1 = t_2^1\}$. (See the Appendix.)

The statements in (8)–(10) do not involve the group-specific effects, and they can therefore be used to construct estimators for β^1, β^2, δ^1 and δ^2 as in Section 2.

3.1.3. Comparing second spells

It is also possible to use two second spells to construct probability statements that are informative about β^2 and δ^2. This does not require the group-specific effect to be the same

[2]It is unlikely that one would assume that $\alpha_i^1 = \alpha_i^2$ without also assuming that $\beta^1 = \beta^2$ and $\delta^1 = \delta^2$. Naturally, the discussion in this section applies to that case as well.

across spells. Let t_1^1, t_2^1, t_1^2 and t_2^2 be arbitrary with $t_1^1 < t_1^2, t_2^1 < t_2^2$ and $|t_1^2 - t_2^2| \leqslant \tau$. Define

$$A = \{T_{1i}^1 = t_1^1, T_{1i}^2 = t_1^2, T_{2i}^1 = t_2^1, T_{2i}^2 > t_2^2\}$$

and

$$B = \{T_{1i}^1 = t_1^1, T_{1i}^2 > t_1^2, T_{2i}^1 = t_2^1, T_{2i}^2 = t_2^2\}.$$

Under the logit Assumption 2a, we then have

$$\mathrm{P}(A|A \cup B, x_{1it_1^2}, x_{2it_2^2}, z_i) = \frac{\exp((x'_{1it_1^2} - x'_{2it_2^2})\beta^2 + \delta^2_{t_1^2 - t_1^1} - \delta^2_{t_2^2 - t_2^1})}{1 + \exp((x'_{1it_1^2} - x'_{2it_2^2})\beta^2 + \delta^2_{t_1^2 - t_1^1} - \delta^2_{t_2^2 - t_2^1})}. \tag{11}$$

Under Assumption 2b

$$\mathrm{P}(A|A \cup B, x_{1it_1^2}, x_{2it_2^2}, z_i) \begin{cases} > \frac{1}{2} & \text{if } (x'_{1it_1^2} - x'_{2it_2^2})\beta^2 + \delta^2_{t_1^2 - t_1^1} - \delta^2_{t_2^2 - t_2^1} > 0, \\ = \frac{1}{2} & \text{if } (x'_{1it_1^2} - x'_{2it_2^2})\beta^2 + \delta^2_{t_1^2 - t_1^1} - \delta^2_{t_2^2 - t_2^1} = 0, \\ < \frac{1}{2} & \text{if } (x'_{1it_1^2} - x'_{2it_2^2})\beta^2 + \delta^2_{t_1^2 - t_1^1} - \delta^2_{t_2^2 - t_2^1} < 0. \end{cases} \tag{12}$$

Finally, under Assumption 2c, and if $t_1^2 - t_1^1 = t_2^2 - t_2^1$,

$$\mathrm{P}(A|A \cup B, x_{1it_1^2}, x_{2it_2^2}, z_i) \begin{cases} > \frac{1}{2} & \text{if } (x'_{1it_1^2} - x'_{2it_2^2})\beta^2 > 0, \\ = \frac{1}{2} & \text{if } (x'_{1it_1^2} - x'_{2it_2^2})\beta^2 = 0, \\ < \frac{1}{2} & \text{if } (x'_{1it_1^2} - x'_{2it_2^2})\beta^2 < 0. \end{cases} \tag{13}$$

Since (11)–(13) do not depend on t_1^1 and t_2^1, the same statements are true if we redefine A and B as $A = \{T_{1i}^2 = t_1^2, T_{2i}^2 > t_2^2\}$ and $B = \{T_{1i}^2 > t_1^2, T_{2i}^2 = t_2^2\}$. As before, this can be used to construct estimators for β^2 and δ^2 without making assumptions on the group-specific effects.

4. An empirical application

In this section, we will use the estimation technique developed above to investigate employee turnover. There are three stylized facts about inter-firm mobility (see Farber, 1999). First, long term employment relationships are common; second, most new jobs end early; and third, the probability of a job ending declines with tenure. The probability of a job separation, however, is generally not equally distributed across individuals and firms. Therefore it is important to control for both individual and firm characteristics. The data set used here is the Integrated Database for Labour Market Research (IDA), which contains information on all employees of all establishments in the private sector in Denmark from 1980 to 2000. Individuals and firms are matched once every year and carry unique identifiers that allow us to follow both individuals and firms over time.

The total number of yearly full time private sector employer–employee matches in the data set is 29,069,419. These are generated by 3,253,312 unique individuals who are working in 477,619 different workplaces. The analysis is conducted on a flow-sample for 5% of the individuals which corresponds to 638,515 observations. The sampling scheme implies that tenure is known for all the employees included in the sample. The average number of employees in a given workplace in a given year is 1.63 with a standard deviation of 2.53. The largest group has 180 members.

The descriptive statistics for the sample used in the analysis are presented in Table 1. Columns one and two present the numbers for women and men, respectively, and the last column shows the numbers for the pooled sample. Women constitute 38.6% of the sample. The three age categories used are below 30 years of age, 30–50 years and above 50 years of age. The largest group is young workers (which is partly caused by the sampling scheme) who account for 46.6% of the individuals. The education level is divided into unskilled (less than 12 years of education), skilled (12–15 years of education) and high-skilled (at least 16 years of education) workers. Skilled workers clearly dominate with a proportion close to 57% (58.3% for men and 54.1% for women). This is a result of the well functioning apprenticeship program and a developed educational market for semiskilled professionals.

Average tenure is 2.41 years with a standard deviation of 3.20. This relatively low number is a result of the flow-sampling scheme that is based on a continuous inflow of newly hired employees and a right censoring in 2000. Hence, the maximum number of years of tenure observed in the sample is 18 years. For the group of employees entering the sample in 1980, 2.59% have employment spells of at least 18 years.

The characteristics of the workplaces included in the sample are presented in the lower part of Table 1. The average (employee-weighted) workplace size is 192. These workplaces have an average payroll per worker in 1980-prices equal to 85,576 Danish Kroners ($\approx \$15,000$). The standard deviation of the payroll measure is 38,434. Finally the distribution of employees across sectors is presented. The largest sector is manufacturing which accounts for 29.1% of the employees.

Since we have discrete time data, the hazard function for employment duration can be characterized by the conditional probability of job separation given a set of explanatory variables. Several studies have shown how individual characteristics such as age, gender, education, marital status and children affect the separation probability, see for example Blau and Kahn (1981), Light and Ureta (1992), Lynch (1972) and Royalty (1998). Others have documented that larger firms and firms with a higher payroll per worker experience lower turnover, see for example Anderson and Meyer (1994). More recently Frederiksen (2004) studied the separation process using employer–employee data, which allowed for effects of both individual and firm characteristics on the job separation process.

Table 2 uses a conventional logit model to estimate the probability of a job separation for men. In column 1, only characteristics of the individual are included. As expected, family related variables such as marriage or cohabitation and having kids reduce the probability of leaving a job significantly. The results also show that the separation rate is declining in age. Finally, men with higher education have lower rate of separations. Column 2 adds information about the workplace. The results show that payroll per worker reduces the separation probability and that a higher variation in pay (standard deviation of the payroll measure) conditional on the payroll level leads to more separations. Furthermore, workplace size has an inverted U-shaped effect on the probability that an employee is leaving the workplace.

In general, controlling for workplace characteristics reduces the magnitude of the coefficients of the individual characteristics but the sign and the significance are preserved. The exception is education. Without controls for firm characteristics, education increases job stability but once the controls are added, skilled workers have higher separation rates

Table 1
Descriptive statistics, 1980–2000

	Women	Men	All
Gender	–	–	0.386
Age less than 30 years	0.491	0.450	0.466
Age 30–50 years	0.377	0.400	0.391
Age above 50 years	0.132	0.150	0.143
Unskilled	0.427	0.375	0.395
Skilled	0.541	0.583	0.567
High skilled	0.032	0.042	0.038
Children	0.345	0.345	0.345
Married/cohabiting	0.458	0.467	0.464
Manufacturing	0.234	0.327	0.291
Primary sector	0.019	0.043	0.033
Electricity, gas and water supply	0.003	0.009	0.007
Construction	0.024	0.132	0.090
Retail and trade	0.310	0.238	0.266
Transportation	0.043	0.073	0.062
Financial	0.202	0.127	0.156
Service	0.166	0.052	0.096
Years of tenure	2.314	2.471	2.411
	(3.133)	(3.241)	(3.201)
Workplace size*	198	208	204
	(63)	(62)	(62)
Workplace size*	186	195	192
(lagged one year)	(61)	(61)	(61)
Payroll per worker*	78,361	90,107	85,576
(1980-prices)	(39,325)	(37,154)	(38,434)
# Observations	246,316	392,199	638,515

Note: Based on a 5% sample. Standard deviations are in parentheses.
*These numbers are employee-weighted.

than both unskilled and highly skilled employees. This suggests that highly skilled employees tend to work in high paying workplaces that in turn have relatively lower turnover on average. Introducing information about the sector of employment (column 3) alters the workplace size coefficients but the rest are insensitive.

The results from the conventional logit models for women are presented in Table 3. The coefficients are generally larger in magnitude than for men, but the relative importance of the explanatory variables is the same as for men. The exception is the coefficient of children, which is smaller for women and statistically insignificant.

Adding firm characteristics has the same effect on the coefficients as for men. The main difference is that for women, the changes in the coefficients for education are not large enough to reverse their signs.

It is clear from the first three columns of Tables 2 and 3 that it is important to include firm-specific variables. This suggests that it is also interesting to allow for unobserved firm characteristics in the way described earlier.

The results of the fixed-effects model (with $\tau = \infty$) are presented in columns 4 and 5 of Tables 2 and 3. The changes in the coefficients suggest that allowing for unobserved

Table 2
Job separation models, men

	Conventional logit model	Conventional logit model	Conventional logit model	Fixed-effects model, $\tau = \max$	Fixed-effects model, $\tau = \max$
Constant	0.110	0.477	0.341		
	(0.020)	(0.021)	(0.023)		
Age less than 30 years	–	–	–	–	–
Age 30–50 years	−0.361	−0.281	−0.286	−0.442	−0.438
	(0.010)	(0.010)	(0.010)	(0.017)	(0.017)
Age more than 50 years	−0.355	−0.277	−0.278	−0.354	−0.348
	(0.014)	(0.014)	(0.014)	(0.022)	(0.022)
Unskilled	–	–	–	–	–
Skilled	−0.018	0.041	0.040	0.022	0.025
	(0.008)	(0.008)	(0.008)	(0.013)	(0.013)
High skilled	−0.150	−0.001	−0.020	−0.081	−0.076
	(0.020)	(0.020)	(0.021)	(0.035)	(0.035)
Children	−0.133	−0.120	−0.116	−0.115	−0.114
	(0.010)	(0.010)	(0.010)	(0.016)	(0.016)
Married/cohabiting	−0.135	−0.105	−0.099	−0.110	−0.110
	(0.011)	(0.011)	(0.011)	(0.016)	(0.016)
Lagged workplace size*		0.078	0.138		0.795
		(0.015)	(0.016)		(0.094)
Lagged workplace size^{2}*		−0.029	−0.039		−0.086
		(0.003)	(0.003)		(0.017)
Payroll per worker**		−0.796	−0.806		−0.362
		(0.014)	(0.015)		(0.059)
Std. dev. of payroll per worker**		0.224	0.247		0.149
		(0.020)	(0.020)		(0.065)
Manufacturing			–		
Primary sector			0.402		
			(0.018)		
Electricity, gas and water supply			−0.268		
			(0.049)		
Construction			0.141		
			(0.012)		
Retail and trade			0.105		
			(0.010)		
Transportation			0.299		
			(0.015)		
Financial			0.257		
			(0.013)		
Service			0.084		
			(0.018)		
Year dummies	YES	YES	YES	YES	YES
Tenure dummies	YES	YES	YES	YES	YES
Log likelihood/objective function	−227,456	−225,269	−224,742	−111,454	−111,159

Note: Based on 392,199 observations.
*Divided by 1000.
**Divided by 100,000.

Table 3
Job separation models, women

	Conventional logit model	Conventional logit model	Conventional logit model	Fixed-effects model, $\tau = \max$	Fixed-effects model, $\tau = \max$
Constant	0.216	0.552	0.528		
	(0.026)	(0.027)	(0.029)		
Age less than 30 years	–	–	–	–	–
Age 30–50 years	−0.575	−0.487	−0.484	−0.571	−0.568
	(0.012)	(0.013)	(0.013)	(0.021)	(0.021)
Age more than 50 years	−0.494	−0.442	−0.438	−0.439	−0.434
	(0.017)	(0.017)	(0.017)	(0.029)	(0.029)
Unskilled	–	–	–	–	–
Skilled	−0.179	−0.069	−0.081	−0.063	−0.060
	(0.009)	(0.010)	(0.010)	(0.017)	(0.017)
High skilled	−0.257	−0.103	−0.073	−0.065	−0.062
	(0.028)	(0.028)	(0.028)	(0.050)	(0.050)
Children	−0.023	−0.001	0.002	−0.007	−0.006
	(0.012)	(0.012)	(0.012)	(0.020)	(0.020)
Married/cohabiting	−0.200	−0.193	−0.197	−0.153	−0.153
	(0.012)	(0.012)	(0.012)	(0.021)	(0.021)
Lagged workplace size*		0.151	0.158		0.429
		(0.019)	(0.020)		(0.112)
Lagged workplace size^{2*}		−0.041	−0.041		−0.031
		(0.004)	(0.004)		(0.018)
Payroll per worker**		−0.893	−0.934		−0.324
		(0.019)	(0.020)		(0.086)
Std. dev. of payroll per worker**		0.214	0.209		0.103
		(0.027)	(0.027)		(0.094)
Manufacturing			–		
Primary sector			0.259		
			(0.033)		
Electricity, gas and water supply			0.024		
			(0.085)		
Construction			−0.058		
			(0.032)		
Retail and trade			0.054		
			(0.013)		
Transportation			0.101		
			(0.025)		
Financial			0.188		
			(0.014)		
Service			−0.054		
			(0.015)		
Year dummies	YES	YES	YES	YES	YES
Tenure dummies	YES	YES	YES	YES	YES
Log likelihood/ objective function	−145,705	−143,785	−143,614	−60,776	−60,695

Note: Based on 246,316 observations.
*Divided by 1000.
**Divided by 100,000.

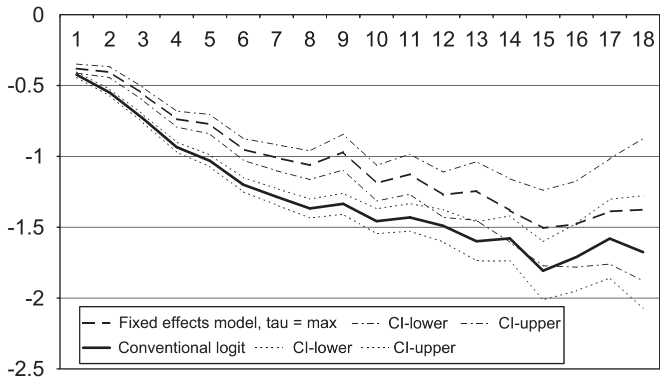

Fig. 1. Coefficients on tenure dummies (men).

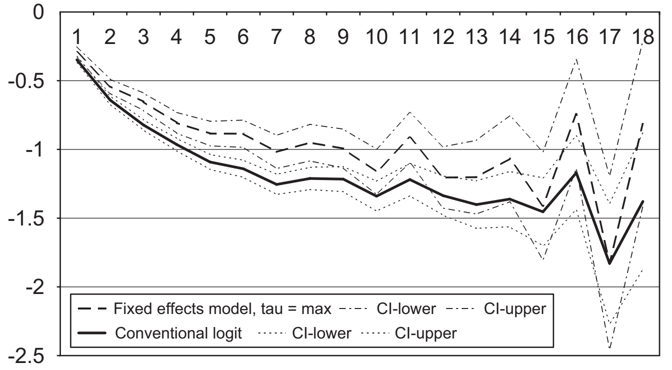

Fig. 2. Coefficients on tenure dummies (women).

firm-specific characteristics is important. A Hausman-type test firmly rejects the hypothesis that the coefficients are the same with or without unobserved firm-specific characteristics (see the first two rows of the last column of Tables 6 and 7). When $\tau = 0$, we test whether the coefficients on the individual characteristics and on the tenure dummies are the same when we use our estimator that allows for unobserved firm-specific effects as when we include firm characteristics (including sector dummies) in a conventional logit model. When $\tau > 0$, we additionally test the equality of the coefficients on time dummies and on time-varying firm–level characteristics. When $\tau = 0$, these are not identified when we allow for unobserved firm-specific effects.[3] The difference between the two versions of the test is that the first implicitly assumes that the conventional logit is asymptotically efficient under the null. It seems reasonable that inferences in the conventional logit model should allow for clustering at the firm level, which would imply that the conventional logit is not asymptotically efficient. The second Hausman-type test statistic is calculated by

[3]The test has 24 degrees of freedom for $\tau = 0$ and 46 for $\tau > 0$.

Table 4
Job separation models, men

	Fixed-effects model				
	$\tau = 0$	$\tau = 1$	$\tau = 5$	$\tau = 10$	$\tau = \max$
Age less than 30 years		–	–	–	–
Age 30–50 years	−0.494	−0.453	−0.451	−0.444	−0.438
	(0.027)	(0.022)	(0.018)	(0.017)	(0.017)
Age more than 50 years	−0.413	−0.374	−0.344	−0.340	−0.348
	(0.035)	(0.029)	(0.024)	(0.023)	(0.022)
Unskilled	–	–	–	–	–
Skilled	0.072	0.054	0.026	0.019	0.025
	(0.021)	(0.018)	(0.014)	(0.013)	(0.013)
High skilled	−0.071	−0.082	−0.062	−0.078	−0.076
	(0.054)	(0.045)	(0.037)	(0.035)	(0.035)
Children	−0.101	−0.099	−0.111	−0.114	−0.114
	(0.025)	(0.020)	(0.017)	(0.016)	(0.016)
Married/cohabiting	−0.130	−0.131	−0.110	−0.107	−0.110
	(0.026)	(0.021)	(0.017)	(0.017)	(0.016)
Lagged workplace size*		0.573	0.850	0.824	0.795
		(0.207)	(0.121)	(0.098)	(0.094)
Lagged workplace size^{2}*		−0.079	−0.083	−0.082	−0.086
		(0.044)	(0.021)	(0.017)	(0.017)
Payroll per worker**		0.198	−0.258	−0.340	−0.362
		(0.112)	(0.069)	(0.062)	(0.059)
Std. dev. of payroll per worker**		−0.293	0.099	0.142	0.149
		(0.130)	(0.081)	(0.069)	(0.065)
Year dummies	NO	YES	YES	YES	YES
Tenure dummies	YES	YES	YES	YES	YES
Objective function	−7998	−24,396	−71,194	−99,829	−111,158

Note: Based on 392,199 observations.
*Divided by 1000.
**Divided by 100,000.

constructing the joint asymptotic distribution of the two estimators assuming independence across firms but allowing for correlations within firms. See Appendix A.3 for details.

A common criticism of the fixed-effect approach is that it makes it hard to estimate marginal effects. This depends on the exact marginal effect of interest. Suppose, for example, that one wants to find the effect of being married on an unmarried man with a separation probability of 10%.[4] For that individual, $(x'_{jit}\beta + \delta_{S_{jit}} + \alpha_i)$ equals -2.197. With the marriage coefficient of -0.110, this implies a fall in the separation probability to 9%. It is tempting to calculate this marginal effect for each model. However, it is not surprising that one would calculate different marginal effects because different sets of additional explanatory variables are kept constant.

[4]Of course, this approach does not allow one to calculate the marginal effect for an individual with given values of the covariates.

Table 5
Job separation models, women

	Fixed-effects model				
	$\tau = 0$	$\tau = 1$	$\tau = 5$	$\tau = 10$	$\tau = \max$
Age less than 30 years		–	–	–	–
Age 30–50 years	−0.617	−0.583	−0.566	−0.566	−0.568
	(0.037)	(0.030)	(0.024)	(0.022)	(0.021)
Age more than 50 years	−0.440	−0.419	−0.413	−0.421	−0.434
	(0.050)	(0.041)	(0.032)	(0.030)	(0.029)
Unskilled	–	–	–	–	–
Skilled	−0.053	−0.045	−0.056	−0.052	−0.060
	(0.029)	(0.024)	(0.019)	(0.018)	(0.017)
High skilled	−0.020	−0.008	−0.045	−0.046	−0.062
	(0.088)	(0.070)	(0.056)	(0.052)	(0.050)
Children	0.026	−0.011	−0.006	−0.006	−0.006
	(0.034)	(0.027)	(0.022)	(0.021)	(0.020)
Married/cohabiting	−0.200	−0.173	−0.166	−0.152	−0.153
	(0.034)	(0.028)	(0.022)	(0.021)	(0.021)
Lagged workplace size*		0.196	0.464	0.466	0.429
		(0.221)	(0.127)	(0.113)	(0.112)
Lagged workplace size^{2*}		0.006	−0.018	−0.031	−0.031
		(0.053)	(0.021)	(0.018)	(0.018)
Payroll per worker**		−0.009	−0.214	−0.268	−0.324
		(0.170)	(0.103)	(0.090)	(0.086)
Std. dev. of payroll per worker**		−0.106	−0.004	0.061	0.103
		(0.179)	(0.110)	(0.098)	(0.094)
Year dummies	NO	YES	YES	YES	YES
Tenure dummies	YES	YES	YES	YES	YES
Objective function	−4,409	−13,351	−39,085	−54,688	−60,695

Note: Based on 246,316 observations.
*Divided by 1,000.
**Divided by 100,000.

Table 6
Test statistics, men

	Fixed-effects model				
	$\tau = 0$	$\tau = 1$	$\tau = 5$	$\tau = 10$	$\tau = \max$
Hausman I	139	1651	2268	2340	2155
Hausman II	140	3944	1712	1889	1705
Wald	1350	1700	2132	2243	2260

Now we turn to the effect of controlling for firm-specific characteristics on the estimates of duration dependence, i.e., the δ's. As discussed in Section 2, a location normalization is needed for parameter identification. We therefore set the δ associated with the shortest tenure (one year) to zero. The estimates of the rest of δ's (along with their pointwise 95%

Table 7
Test statistics, women

	Fixed-effects model				
	$\tau = 0$	$\tau = 1$	$\tau = 5$	$\tau = 10$	$\tau = \max$
Hausman I	47	834	1027	1144	1071
Hausman II	46	1284	251	625	652
Wald	631	818	1174	1363	1365

confidence intervals) are plotted in Fig. 1 (for men) and Fig. 2 (for women). For both conventional logit and fixed-effect models, all coefficients are negative and they are decreasing as a function of duration, which indicates negative duration dependence. However, the estimates from the fixed-effect models are uniformly smaller in magnitude than those from the conventional logit model, which suggests a lesser degree of duration dependence once firm-specific effects are controlled for.

It is evident from the figures that the duration dependence coefficients are jointly significantly different from zero. This is confirmed by the Wald test presented in the last rows of Tables 6 and 7.

The results based on $\tau = \infty$ presented in the last column of Tables 2 and 3 assume that there is no feedback from one worker's dependent variable to the future explanatory variables of other workers in the same firm. This might not be reasonable for time-varying firm-level explanatory variables. These are presumably chosen by the firm taking into account all the relevant information including past turnovers of its other workers. Tables 4 and 5 present results for different values of τ where one can think of τ as the time it takes for the firm to adjust its aggregate variable. Note when $\tau = 0$, we cannot identify the effect of firm-level explanatory variables because we implicitly allow for the unobserved firm-specific characteristics to be time varying. Not surprisingly, the point estimates are sensitive to choices of τ. However, the coefficients on individual characteristics are less sensitive than are the coefficients on the firm-level explanatory variables. This is what one would expect since individual-specific variables are less likely to be subject to feedback. Tables 6 and 7 present the Wald- and Hausman-type tests discussed earlier for different values of τ.

5. Conclusions

This paper considers a discrete choice/duration model in which the dynamics is handled by using the duration in the current state as a covariate. The main contribution is to propose estimators that allow for group-specific effect in parametric and semiparametric versions of the model. This is relevant in many empirical settings where one observes individuals that are grouped geographically, by household, by employer, etc. On the other hand, there are also many situations in which one would want to use the models considered here in applications where the grouping results from multiple spells for the same individual. The approaches discussed in this paper do not automatically apply in that case. The reason is that when one observes consecutive spells for the same individual, the timing of the second spell (and hence the covariates for the second spell) will in general depend on the length of the first spell. This will violate the assumptions made in this paper. Investigating methods for dealing with that case is an interesting topic for future research.

Acknowledgment

This research was supported by the National Science Foundation, the Gregory C. Chow Econometric Research Program at Princeton University, and the Danish National Research Foundation (through CAM at The University of Copenhagen and CCP at the Aarhus School of Business). We would like to thank Dante Amengual, Tiemen Woutersen and a number of seminar participants for their suggestions. Part of the material presented here was distributed as "Estimation of Discrete Time Duration Models with Grouped Data". That paper has been retired.

Appendix A

A.1. Derivation of Lemma 1

Let t_1 and t_2 be arbitrary with $|t_1 - t_2| \leqslant \tau$, and recall that z_i denotes the set of predetermined variables for group i at the beginning of the sample.

Consider the two events $A = \{T_{1i} = t_1, T_{2i} > t_2\}$ and $B = \{T_{1i} > t_1, T_{2i} = t_2\}$. Notationally, it will be convenient to distinguish between the case where $t_1 = t_2$ and the case where $t_1 \neq t_2$. In the latter case there is no loss of generality in assuming that $t_1 < t_2$,

$$
\begin{aligned}
&\mathrm{P}(A, \{x_{1it}\}_{t=2}^{t_1}, \{x_{2it}\}_{t=2}^{t_2} | z_i)\\
&\quad = P_1(y_{1i1} = 0, y_{2i1} = 0 | z_i)\\
&\qquad \times p_2(x_{1i2}, x_{2i2} | z_i, y_{1i1} = 0, y_{2i1} = 0)\\
&\qquad \times \cdots\\
&\qquad \times \cdots\\
&\qquad \times P_{t_1}(y_{1it_1} = 1, y_{2it_1} = 0 | z_i, \{x_{1is}, x_{2is}\}_{s \leqslant t_1}, \{y_{1is} = 0, y_{2is} = 0\}_{s < t_1})\\
&\qquad \times p_{t_1+1}(x_{2it_1+1} | z_i, \{x_{1is}, x_{2is}\}_{s \leqslant t_1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leqslant t_1})\\
&\qquad \times P_{t_1+1}(y_{2it_1+1} = 0 | z_i, \{x_{1is}\}_{s \leqslant t_1}, \{x_{2is}\}_{s \leqslant t_1+1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leqslant t_1})\\
&\qquad \times p_{t_1+2}(x_{2it_1+2} | z_i, \{x_{1is}\}_{s \leqslant t_1}, \{x_{2is}\}_{s \leqslant t_1+1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leqslant t_1+1})\\
&\qquad \times P_{t_1+2}(y_{2it_1+2} = 0 | z_i, \{x_{1is}\}_{s \leqslant t_1}, \{x_{2is}\}_{s \leqslant t_1+2}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leqslant t_1+1})\\
&\qquad \times \cdots\\
&\qquad \times \cdots\\
&\qquad \times p_{t_2}(x_{2it_2} | z_i, \{x_{1is}\}_{s \leqslant t_1}, \{x_{2is}\}_{s \leqslant t_2-1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leqslant t_2-1})\\
&\qquad \times P_{t_2}(y_{2it_2} = 0 | z_i, \{x_{1is}\}_{s \leqslant t_1}, \{x_{2is}\}_{s \leqslant t_2}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leqslant t_2-1})
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{P}(B, \{x_{1it}\}_{t=2}^{t_1}, \{x_{2it}\}_{t=2}^{t_2} | z_i)\\
&\quad = P_1(y_{1i1} = 0, y_{2i1} = 0 | z_i)\\
&\qquad \times p_2(x_{1i2}, x_{2i2} | z_i, y_{1i1} = 0, y_{2i1} = 0)
\end{aligned}
$$

$$
\begin{aligned}
&\times\cdots\\
&\times\cdots\\
&\times P_{t_1}(y_{1it_1}=0,y_{2it_1}=0|z_i,\{x_{1is},x_{2is}\}_{s\leqslant t_1},\{y_{1is}=0,y_{2is}=0\}_{s<t_1})\\
&\times p_{t_1+1}(x_{2it_1+1}|z_i,\{x_{1is},x_{2is}\}_{s\leqslant t_1},\{y_{1is}=0\}_{s\leqslant t_1},\{y_{2is}=0\}_{s\leqslant t_1})\\
&\times P_{t_1+1}(y_{2it_1+1}=0|z_i,\{x_{1is}\}_{s\leqslant t_1},\{x_{2is}\}_{s\leqslant t_1+1},\{y_{1is}=0\}_{s\leqslant t_1},\{y_{2is}=0\}_{s\leqslant t_1})\\
&\times p_{t_1+2}(x_{2it_1+2}|z_i,\{x_{1is}\}_{s\leqslant t_1},\{x_{2is}\}_{s\leqslant t_1+1},\{y_{1is}=0\}_{s\leqslant t_1},\{y_{2is}=0\}_{s\leqslant t_1+1})\\
&\times P_{t_1+2}(y_{2it_1+2}=0|z_i,\{x_{1is}\}_{s\leqslant t_1},\{x_{2is}\}_{s\leqslant t_1+2},\{y_{1is}=0\}_{s\leqslant t_1},\{y_{2is}=0\}_{s\leqslant t_1+1})\\
&\times\cdots\\
&\times\cdots\\
&\times p_{t_2}(x_{2it_2}|z_i,\{x_{1is}\}_{s\leqslant t_1},\{x_{2is}\}_{s\leqslant t_2-1},\{y_{1is}=0\}_{s\leqslant t_1},\{y_{2is}=0\}_{s\leqslant t_2-1})\\
&\times P_{t_2}(y_{2it_2}=1|z_i,\{x_{1is}\}_{s\leqslant t_1},\{x_{2is}\}_{s\leqslant t_2},\{y_{1is}=0\}_{s\leqslant t_1},\{y_{2is}=0\}_{s\leqslant t_2-1}).
\end{aligned}
$$

The case where $t_1=t_2$ is dealt with in the same way except that one calculates $\mathrm{P}(A,\{x_{1it}\}_{t=2}^{t_1},\ \{x_{2it}\}_{t=2}^{t_1}|z_i)$ and $\mathrm{P}(B,\{x_{1it}\}_{t=2}^{t_1},\{x_{2it}\}_{t=2}^{t_1}|z_i)$,

$$
\begin{aligned}
&\mathrm{P}(A,\{x_{1it}\}_{t=2}^{t_1},\{x_{2it}\}_{t=2}^{t_2}|z_i)\\
&\quad=P_1(y_{1i1}=0,y_{2i1}=0|z_i)\\
&\quad\times p_2(x_{1i2},x_{2i2}|z_i,y_{1i1}=0,y_{2i1}=0)\\
&\quad\times\cdots\\
&\quad\times\cdots\\
&\quad\times P_{t_1}(y_{1it_1}=1,y_{2it_1}=0|z_i,\{x_{1is},x_{2is}\}_{s\leqslant t_1},\{y_{1is}=0,y_{2is}=0\}_{s<t_1})
\end{aligned}
$$

and similarly for $\mathrm{P}(B,\{x_{1it}\}_{t=2}^{t_1},\{x_{2it}\}_{t=2}^{t_1}|z_i)$.

Either way one concludes that

$$
\begin{aligned}
\mathrm{P}(A|A\cup B,\{x_{1it}\}_{t=2}^{t_1},\{x_{2it}\}_{t=2}^{t_2},z_i)&=\mathrm{P}(A,\{x_{1it}\}_{t=2}^{t_1},\{x_{2it}\}_{t=2}^{t_2}|A\cup B,\{x_{1it}\}_{t=2}^{t_1},\{x_{2it}\}_{t=2}^{t_2},z_i)\\
&=\frac{a_1}{a_1+a_2},
\end{aligned}
$$

where

$$
\begin{aligned}
a_1=\;&P_{t_1}(y_{1it_1}=1,y_{2it_1}=0|z_i,\{x_{1is},x_{2is}\}_{s\leqslant t_1},\{y_{1is}=0,y_{2is}=0\}_{s<t_1})\\
&\times P_{t_2}(y_{2it_2}=0|z_i,\{x_{1is}\}_{s\leqslant t_1},\{x_{2is}\}_{s\leqslant t_2},\{y_{1is}=0\}_{s<t_1},y_{1it_1}=1,\{y_{2is}=0\}_{s\leqslant t_2-1}),
\end{aligned}
$$

$$
\begin{aligned}
a_2=\;&P_{t_1}(y_{1it_1}=0,y_{2it_1}=0|z_i,\{x_{1is},x_{2is}\}_{s\leqslant t_1},\{y_{1is}=0,y_{2is}=0\}_{s<t_1})\\
&\times P_{t_1}(y_{2it_2}=1|z_i,\{x_{1is}\}_{s\leqslant t_1},\{x_{2is}\}_{s\leqslant t_2},\{y_{1is}=0\}_{s\leqslant t_1},\{y_{2is}=0\}_{s\leqslant t_2-1}).
\end{aligned}
$$

Under Assumptions 2a and 2b

$$
a_1=F(x'_{1it_1}\beta+\delta_{t_1+S_{1i1}}+\alpha_i)\cdot(1-F(x'_{2it_1}\beta+\delta_{t_1+S_{2i1}}+\alpha_i))\cdot(1-F(x'_{2it_2}\beta+\delta_{t_2+S_{2i1}}+\alpha_i))
$$

and

$$
a_2=(1-F(x'_{1it_1}\beta+\delta_{t_1+S_{1i1}}+\alpha_i))\cdot(1-F(x'_{2it_1}\beta+\delta_{t_1+S_{2i1}}+\alpha_i))\cdot F(x'_{2it_2}\beta+\delta_{t_2+S_{2i1}}+\alpha_i)
$$

so

$$P(A|A \cup B, \{x_{1it}\}_{t=1}^{t_1}, \{x_{2it}\}_{t=1}^{t_2}, z_i) = \frac{c_1}{c_2},$$

where

$$c_1 = F(x'_{1it_1}\beta + \delta_{t_1+S_{1i1}} + \alpha_i) \cdot (1 - F(x'_{2it_2}\beta + \delta_{t_2+S_{2i1}} + \alpha_i))$$

and

$$c_2 = F(x'_{1it_1}\beta + \delta_{t_1+S_{1i1}} + \alpha_i) \cdot (1 - F(x'_{2it_2}\beta + \delta_{t_2+S_{2i1}} + \alpha_i)) \\ + (1 - F(x'_{1it_1}\beta + \delta_{t_1+S_{1i1}} + \alpha_i)) \cdot F(x'_{2it_2}\beta + \delta_{t_2+S_{2i1}} + \alpha_i).$$

This implies that

$$P(A|A \cup B, x_{1it_1}, x_{2it_2}, z_i) = \frac{c_1}{c_2}.$$

Under Assumption 2a, F is the logistic CDF and

$$P(A|A \cup B, x_{1it_1}, x_{2it_2}, z_i) = \frac{\exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))}{1 + \exp((x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))}.$$

Under Assumption 2b

$$\frac{P(A|x_{1it_1}, x_{2it_2}, z_i)}{P(B|x_{1it_1}, x_{2it_2}, z_i)} = \frac{F(x'_{1it_1}\beta + \delta_{t_1+S_{1i1}} + \alpha_i)}{F(x'_{2it_2}\beta + \delta_{t_2+S_{2i1}} + \alpha_i)} \cdot \frac{1 - F(x'_{2it_2}\beta + \delta_{t_2+S_{2i1}} + \alpha_i)}{1 - F(x'_{1it_1}\beta + \delta_{t_1+S_{1i1}} + \alpha_i)}$$

and therefore

$$P(A|A \cup B, x_{1it_1}, x_{2it_2}, z_i) \begin{cases} > \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) > 0, \\ = \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) = 0, \\ < \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) < 0. \end{cases}$$

Finally, under Assumption 2c

$$a_1 = F_{t_1+S_{1i1}}(x'_{1it_1}\beta + \alpha_i) \cdot (1 - F_{t_1+S_{2i1}}(x'_{2it_1}\beta + \alpha_i)) \cdot (1 - F_{t_2+S_{2i1}}(x'_{2it_2}\beta + \alpha_i))$$

and

$$a_2 = (1 - F_{t_1+S_{1i1}}(x'_{1it_1}\beta + \alpha_i)) \cdot (1 - F_{t_1+S_{2i1}}(x'_{2it_1}\beta + \alpha_i)) \cdot F_{t_2+S_{2i1}}(x'_{2it_2}\beta + \alpha_i)$$

so if $t_1 + S_{1i1} = t_2 + S_{2i1}$,

$$P\big(A|A \cup B, x_{1it_1}, x_{2it_2}, z_i\big) \begin{cases} > \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta > 0, \\ = \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta = 0, \\ < \frac{1}{2} & \text{if } (x_{1it_1} - x_{2it_2})'\beta < 0. \end{cases}$$

A.2. Derivation of results with multiple spells

This section derives the main claims of Section 3.

We will consider three types of events (with corresponding contribution to the objective function). For each of those types of events there are a number of special cases depending on the ordering of t_1^1, t_1^2, t_2^1 and t_2^2 defined below. However, the basic structures of the calculations are the same throughout.

A.2.1. Comparing first spells

One can use the first spells of individuals i_1 and i_2 to construct conditional probability statements like the ones in the previous section.

A.2.2. Comparing first spells to second spells

Let t_1^1, t_1^2 and t_2^1 be arbitrary with $t_1^1 < t_1^2$ and $|t_1^2 - t_2^1| \leqslant \tau$, and let z_i denote the set of predetermined variables for group i at the beginning of the sample.

Consider the two events $A = \{T_{1i}^1 = t_1^1, T_{1i}^2 = t_1^2, T_{2i}^1 > t_2^1\}$ and $B = \{T_{1i}^1 = t_1^1, T_{1i}^2 > t_1^2, T_{2i}^1 = t_2^1\}$. We will consider three cases based on the ordering of t_1^1, t_2^1, and t_1^2. The calculation below is for the case where $1 < t_1^1 < t_2^1 < t_1^2$ (the other cases follow in exactly the same manner),

$$
\begin{aligned}
&\mathrm{P}(A, \{x_{1it}\}_{t=2}^{t_1^1+t_1^2}, \{x_{2it}\}_{t=2}^{t_2^1} | z_i) \\
&\quad = P_1(y_{1i1}^1 = 0, y_{2i1}^1 = 0 | z_i) \\
&\qquad \times p_2(x_{1i2}, x_{2i2} | z_i, y_{1i1}^1 = 0, y_{2i1}^1 = 0) \\
&\qquad \times \cdots \\
&\qquad \times \cdots \\
&\qquad \times P_{t_1^1}(y_{1it_1^1}^1 = 1, y_{2it_1^1}^1 = 0 | z_i, \{x_{1is}, x_{2is}\}_{s \leqslant t_1^1}, \{y_{1is}^1 = 0, y_{2is}^1 = 0\}_{s < t_1^1}) \\
&\qquad \times p_{t_1^1+1}(x_{1it_1^1+1}, x_{2it_1^1+1} | z_i, \{x_{1is}, x_{2is}\}_{s \leqslant t_1^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{2is}^1 = 0\}_{s \leqslant t_1^1}) \\
&\qquad \times P_{t_1^1+1}(y_{1it_1^1+1}^2 = 0, y_{2it_1^1+1}^1 = 0 | z_i, \{x_{1is}\}_{s \leqslant t_1^1+1}, \{x_{2is}\}_{s \leqslant t_1^1+1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, \\
&\qquad\quad y_{1it_1}^1 = 1, \{y_{2is}^1 = 0\}_{s \leqslant t_1^1}) \\
&\qquad \times p_{t_1^1+2}(x_{1it_1+2}, x_{2it_1+2} | z_i, \{x_{1is}\}_{s \leqslant t_1^1+1}, \{x_{2is}\}_{s \leqslant t_1^1+1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1}^1 = 1, \\
&\qquad\quad y_{1it_1^1+1}^2 = 0, \{y_{2is}^1 = 0\}_{s \leqslant t_1^1+1}) \\
&\qquad \times P_{t_1^1+2}(y_{1it_1+2}^2 = 0, y_{2it_1+2}^1 = 0 | z_i, \{x_{1is}\}_{s \leqslant t_1^1+2}, \{x_{2is}\}_{s \leqslant t_1^1+2}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \\
&\qquad\quad y_{1it_1+1}^2 = 0, \{y_{2is}^1 = 0\}_{s \leqslant t_1^1+1}) \\
&\qquad \times \cdots \\
&\qquad \times \cdots \\
&\qquad \times p_{t_2^1}(x_{1it_2^1}, x_{2it_2^1} | z_i, \{x_{1is}\}_{s \leqslant t_2^1-1}, \{x_{2is}\}_{s \leqslant t_2^1-1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_2^1-1}, \\
&\qquad\quad \{y_{2is}^1 = 0\}_{s \leqslant t_2^1-1}) \\
&\qquad \times P_{t_2^1}(y_{1it_2^1}^2 = 0, y_{2it_2^1}^1 = 0 | z_i, \{x_{1is}\}_{s \leqslant t_2^1}, \{x_{2is}\}_{s \leqslant t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \\
&\qquad\quad \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_2^1-1}, \{y_{2is}^1 = 0\}_{s \leqslant t_2^1-1}) \\
&\qquad \times p_{t_2^1+1}(x_{1it_2^1+1} | z_i, \{x_{1is}\}_{s \leqslant t_2^1}, \{x_{2is}\}_{s \leqslant t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \\
&\qquad\quad \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_2^1}, \{y_{2is}^1 = 0\}_{s \leqslant t_2^1})
\end{aligned}
$$

$$\begin{aligned}
&\times P_{t_2^1+1}(y^2_{1it_2^1+1}=0|z_i,\{x_{1is}\}_{s\leqslant t_2^1+1},\{x_{2is}\}_{s\leqslant t_2^1},\{y^1_{1is}=0\}_{s<t_1^1},y^1_{1it_1^1}=1,\\
&\quad \{y^2_{1is}=0\}_{s=t_1^1+1}^{t_2^1},\{y^1_{2is}=0\}_{s\leqslant t_2^1})\\
&\times\cdots\\
&\times\cdots\\
&\times p_{t_1^2}(x_{1it_1^2}|z_i,\{x_{1is}\}_{s\leqslant t_1^2-1},\{x_{2is}\}_{s\leqslant t_2^1},\{y^1_{1is}=0\}_{s<t_1^1},y^1_{1it_1^1}=1,\\
&\quad \{y^2_{1is}=0\}_{s=t_1^1+1}^{t_1^2-1},\{y^1_{2is}=0\}_{s\leqslant t_2^1})\\
&\times P_{t_1^2}(y^2_{1it_1^2}=1|z_i,\{x_{1is}\}_{s\leqslant t_1^2},\{x_{2is}\}_{s\leqslant t_2^1},\{y^1_{1is}=0\}_{s<t_1^1},y^1_{1it_1^1}=1,\\
&\quad \{y^2_{1is}=0\}_{s=t_1^1+1}^{t_1^2-1},\{y^1_{2is}=0\}_{s\leqslant t_2^1}).
\end{aligned}$$

$\mathrm{P}(B,\{x_{1it}\}_{t=2}^{t_1^1+t_1^2},\{x_{2it}\}_{t=2}^{t_2^1}|z_i)$ is derived in exactly the same manner. We therefore have

$$\begin{aligned}
&\mathrm{P}(A|A\cup B,\{x_{1it}\}_{t=2}^{t_1^1+t_1^2},\{x_{2it}\}_{t=2}^{t_2^1},z_i)\\
&\quad=\mathrm{P}(A,\{x_{1it}\}_{t=2}^{t_1^1+t_1^2},\{x_{2it}\}_{t=2}^{t_2^1}|A\cup B,\{x_{1it}\}_{t=2}^{t_1^1+t_1^2},\{x_{2it}\}_{t=2}^{t_2^1},z_i)\\
&\quad=\frac{a_1}{a_1+a_2},
\end{aligned}$$

where

$$\begin{aligned}
a_1&=P_{t_2^1}(y^2_{1it_2^1}=0,y^1_{2it_2^1}=0|z_i,\{x_{1is}\}_{s\leqslant t_2^1},\{x_{2is}\}_{s\leqslant t_2^1},\{y^1_{1is}=0\}_{s<t_1^1},y^1_{1it_1^1}=1,\\
&\quad \{y^2_{1is}=0\}_{s=t_1^1+1}^{t_2^1-1},\{y^1_{2is}=0\}_{s\leqslant t_2^1-1})\\
&\quad\times P_{t_1^2}(y^2_{1it_1^2}=1|z_i,\{x_{1is}\}_{s\leqslant t_1^2},\{x_{2is}\}_{s\leqslant t_2^1},\{y^1_{1is}=0\}_{s<t_1^1},y^1_{1it_1^1}=1,\\
&\quad \{y^2_{1is}=0\}_{s=t_1^1+1}^{t_1^2-1},\{y^1_{2is}=0\}_{s\leqslant t_2^1})\\
&=(1-F_{t_2^1}(x'_{1it_2^1}\beta^1+\delta^1_{t_2^1-t_1^1}+\alpha_i^1))\cdot(1-F_{t_2^1}(x'_{2it_2^1}\beta^1+\delta^1_{t_2^1+S_{2i1}}+\alpha_i^1))\\
&\quad\times F_{t_1^2}(x'_{1it_1^2}\beta^2+\delta^2_{t_1^2-t_1^1}+\alpha_i^2),
\end{aligned}$$

$$\begin{aligned}
a_2&=P_{t_2^1}(y^2_{1it_2^1}=0,y^1_{2it_2^1}=1|z_i,\{x_{1is}\}_{s\leqslant t_2^1},\{x_{2is}\}_{s\leqslant t_2^1},\{y^1_{1is}=0\}_{s<t_1^1},y^1_{1it_1^1}=1,\\
&\quad \{y^2_{1is}=0\}_{s=t_1^1+1}^{t_2^1-1},\{y^1_{2is}=0\}_{s\leqslant t_2^1-1})\\
&\quad\times P_{t_1^2}(y^2_{1it_1^2}=0|z_i,\{x_{1is}\}_{s\leqslant t_1^2},\{x_{2is}\}_{s\leqslant t_2^1},\{y^1_{1is}=0\}_{s<t_1^1},y^1_{1it_1^1}=1,\\
&\quad \{y^2_{1is}=0\}_{s=t_1^1+1}^{t_1^2-1},\{y^1_{2is}=0\}_{s<t_2^1},y^1_{2it_2^1}=1)\\
&=(1-F_{t_2^1}(x'_{1it_2^1}\beta^1+\delta^1_{t_2^1-t_1^1}+\alpha_i^1))\cdot F_{t_2^1}(x'_{2it_2^1}\beta^1+\delta^1_{t_2^1+S_{2i1}}+\alpha_i^1)\\
&\quad\times(1-F_{t_1^2}(x'_{1it_1^2}\beta^2+\delta^2_{t_1^2-t_1^1}+\alpha_i^2))
\end{aligned}$$

so

$$
\begin{aligned}
&\mathrm{P}(A|A\cup B,\{x_{1it}\}_{t=1}^{t_1},\{x_{2it}\}_{t=1}^{t_2},z_i)\\
&=\frac{\dfrac{(1-F_{t_2^1}(x'_{2it_2^1}\beta^1+\delta^1_{t_2^1+S_{2i1}}+\alpha_i^1))\cdot F_{t_1^2}(x'_{1it_1^2}\beta^2+\delta^2_{t_1^2-t_1^1}+\alpha_i^2)}{F_{t_2^1}(x'_{2it_2^1}\beta^1+\delta^1_{t_2^1+S_{2i1}}+\alpha_i^1)\cdot(1-F_{t_1^2}(x'_{1it_1^2}\beta^2+\delta^2_{t_1^2-t_1^1}+\alpha_i^2))}}{1+\dfrac{(1-F_{t_2^1}(x'_{2it_2^1}\beta^1+\delta^1_{t_2^1+S_{2i1}}+\alpha_i^1))\cdot F_{t_1^2}(x'_{1it_1^2}\beta^2+\delta^2_{t_1^2-t_1^1}+\alpha_i^2)}{F_{t_2^1}(x'_{2it_2^1}\beta^1+\delta^1_{t_2^1+S_{2i1}}+\alpha_i^1)\cdot(1-F_{t_1^2}(x'_{1it_1^2}\beta^2+\delta^2_{t_1^2-t_1^1}+\alpha_i^2))}}. \qquad (A.1)
\end{aligned}
$$

Unless $\alpha_i^1=\alpha_i^2$ this will not lead to expressions that can be used to make inference about β and the duration dependence parameters without additional assumptions on the group-specific effects α_i^1 and α_i^2. Of course, there are many cases in which it would be reasonable to assume that the model (including the group-specific effects) are constant from spell to spell. In that case (A.1) implies that under Assumption 2a,[5]

$$
\mathrm{P}(A|A\cup B,x_{1it_1^2},x_{2it_2^1},z_i)=\frac{\exp(x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1+\delta^2_{t_1^2-t_1^1}-\delta^1_{t_2^1+S_{2i1}})}{1+\exp(x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1+\delta^2_{t_1^2-t_1^1}-\delta^1_{t_2^1+S_{2i1}})} \qquad (A.2)
$$

and under Assumption 2b

$$
\mathrm{P}(A|A\cup B,x_{1it_1},x_{2it_2},z_i)\begin{cases}>\frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1+\delta^2_{t_1^2-t_1^1}-\delta^1_{t_2^1+S_{2i1}}>0,\\ =\frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1+\delta^2_{t_1^2-t_1^1}-\delta^1_{t_2^1+S_{2i1}}=0,\\ <\frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1+\delta^2_{t_1^2-t_1^1}-\delta^1_{t_2^1+S_{2i1}}<0.\end{cases} \qquad (A.3)
$$

Finally, under Assumption 2c, and if $t_1^2-t_1^1=t_2^1+S_{2i1}$,

$$
\mathrm{P}(A|A\cup B,x_{1it},x_{2it},z_i)\begin{cases}>\frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1>0,\\ =\frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1=0,\\ <\frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2-x'_{2it_2^1}\beta^1<0.\end{cases} \qquad (A.4)
$$

Since (A.2)–(A.4) do not depend on t_1^1 and t_2^1, the same statements are true if we redefine A and B as $\tilde{A}=\{T_{1i}^2=t_1^2,T_{2i}^1>t_2^1\}$ and $\tilde{B}=\{T_{1i}^2>t_1^2,T_{2i}^1=t_2^1\}$. To see why, note that

$$
\begin{aligned}
\mathrm{P}(A|A\cup B,x_{1it},x_{2it},z_i)&=\mathrm{P}(\tilde{A}|\tilde{A}\cup\tilde{B},x_{1it},x_{2it},z_i,T_{1i}^1=t_1^1)\\
&=\mathrm{P}(\tilde{A}|\tilde{A}\cup\tilde{B},x_{1it},x_{2it},z_i)
\end{aligned}
$$

(since the left-hand side does not depend on t_1^1).

A.2.3. Comparing two second spells

We next turn to the case where we compare the duration of the second spell for two individuals. Let t_1^1, t_1^2, t_2^1 and t_2^2 be arbitrary with $t_1^1<t_1^2$, $t_2^1<t_2^2$ and $|t_1^2-t_2^2|\leqslant\tau$, and recall

[5] In this case it would be reasonable to impose $\beta^1=\beta^2$ and $\delta_\tau^1=\delta_\tau^2$. This would further change the notation, so we do not impose this restriction.

that z_i denotes the set of predetermined variables for group i at the beginning of the sample.

Consider the two events $A = \{T^1_{1i} = t^1_1, T^2_{1i} = t^2_1, T^1_{2i} = t^1_2, T^2_{2i} > t^2_2\}$ and $B = \{T^1_{1i} = t^1_1, T^2_{1i} > t^2_1, T^1_{2i} = t^1_2, T^2_{2i} = t^2_2\}$. Mimicking the calculations above we find that under Assumption 2a

$$\mathrm{P}(A|A \cup B, x_{1it^2_1}, x_{2it^2_2}, z_i) = \frac{\exp((x'_{1it^2_1} - x'_{2it^2_2})\beta^2 + \delta^2_{t^2_1 - t^1_1} - \delta^2_{t^2_2 - t^1_2})}{1 + \exp((x'_{1it^2_1} - x'_{2it^2_2})\beta^2 + \delta^2_{t^2_1 - t^1_1} - \delta^2_{t^2_2 - t^1_2})} \tag{A.5}$$

and under Assumption 2b

$$\mathrm{P}(A|A \cup B, x_{1it^2_1}, x_{2it^2_2}, z_i)\begin{cases} > \frac{1}{2} & \text{if } (x'_{1it^2_1} - x'_{2it^2_2})\beta^2 + \delta^2_{t^2_1 - t^1_1} - \delta^2_{t^2_2 - t^1_2} > 0, \\ = \frac{1}{2} & \text{if } (x'_{1it^2_1} - x'_{2it^2_2})\beta^2 + \delta^2_{t^2_1 - t^1_1} - \delta^2_{t^2_2 - t^1_2} = 0, \\ < \frac{1}{2} & \text{if } (x'_{1it^2_1} - x'_{2it^2_2})\beta^2 + \delta^2_{t^2_1 - t^1_1} - \delta^2_{t^2_2 - t^1_2} < 0. \end{cases} \tag{A.6}$$

Finally, under Assumption 2c, and if $t^2_1 - t^1_1 = t^2_2 - t^1_2$,

$$\mathrm{P}(A|A \cup B, x_{1it^2_1}, x_{2it^2_2}, z_i)\begin{cases} > \frac{1}{2} & \text{if } (x'_{1it^2_1} - x'_{2it^2_2})\beta^2 > 0, \\ = \frac{1}{2} & \text{if } (x'_{1it^2_1} - x'_{2it^2_2})\beta^2 = 0, \\ < \frac{1}{2} & \text{if } (x'_{1it^2_1} - x'_{2it^2_2})\beta^2 < 0. \end{cases} \tag{A.7}$$

Since (A.5)–(A.7) do not depend on t^1_1 and t^1_2, the same statements are true if we redefine A and B as $A = \{T^2_{1i} = t^2_1, T^2_{2i} > t^2_2\}$ and $B = \{T^2_{1i} > t^2_1, T^2_{2i} = t^2_2\}$.

A.3. Calculating the "Hausman"–type test under clustering

The conventional logit estimator satisfies the usual asymptotic linearity

$$\begin{aligned}\sqrt{n}(\hat{\beta}_{\text{logit}} - \beta) &= \sqrt{n}\left(\sum_{i,j} q''_{ji}(\beta)\right)^{-1} \sum_{i,j} q'_{ji}(\beta) \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{J_i} q''_{ji}(\beta)\right)\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{j=1}^{J_i} q'_{ji}(\beta)\right) + \mathrm{o_p}(1),\end{aligned}$$

where q_{ji} is the contribution to the log–likelihood function for a given individual, j, in group i. Using standard arguments, our estimator (under Assumption 2a) can be written as

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i-1}\sum_{k=2}^{J_i} r''_{jki}(\beta)\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{j=1}^{J_i-1}\sum_{k=2}^{J_i} r'_{jki}(\beta)\right) + \mathrm{o_p}(1),$$

where r_{jki} is the contribution to the objective function for a pair of individuals, j and k, in group i. The joint asymptotic distribution of $\hat{\beta}_{\text{logit}}$ and $\hat{\beta}$ is then obtained from

$$\sqrt{n}\left(\begin{pmatrix} \hat{\beta}_{\text{logit}} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} \beta \\ \beta \end{pmatrix}\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \sum_{j=1}^{J_i} q''_{ji}(\beta) & 0 \\ 0 & \sum_{j=1}^{J_i-1}\sum_{k=2}^{J_i} r''_{jki}(\beta) \end{pmatrix}\right)^{-1} \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \sum_{j=1}^{J_i} q'_{ji}(\beta) \\ \sum_{j=1}^{J_i-1}\sum_{k=2}^{J_i} r'_{jki}(\beta) \end{pmatrix} + o_p(1).$$

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

Andersen, E.B., 1970. Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society, Series B 32, 283–301.

Anderson, P.M., Meyer, B.D., 1994. The extent and consequences of job turnover. Brookings Papers on Economic Activity: Microeconomics, 177–248.

Blau, F.D., Kahn, L.M., 1981. Race and sex differences in quits by young workers. Industrial and Labour Relations Review 34, 563–577.

Chamberlain, G., 1985. Heterogeneity, omitted variable bias, and duration dependence. In: Heckman, J.J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data. Econometric Society Monographs Series, vol. 10. Cambridge University Press, Cambridge, New York, Sydney, pp. 3–38.

Clayton, D., Cuzick, J., 1985. Multivariate generalizations of the proportional hazards model. Journal of the Royal Statistical Society 148, 82–117.

Cox, D.R., 1958. The regression analysis of binary sequences. Journal of the Royal Statistical Society 20, 215–242.

Cox, D. R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society Series B-Statistical Methodology, 34.

Cox, D.R., 1975. Partial likelihood. Biometrika, 62.

D'Adddio, A.C., Honoré, B.E., 2006. Duration Dependence and Time Varying Variables in Discrete Time Duration Models. Princeton University, Princeton, NJ.

Farber, H.S., 1999. Mobility and stability: the dynamics of job change in: labour markets. In: Ashenfelter, O., Card, D. (Eds.), The Handbook of Labour Economics. Handbooks in Economics, vol. 3. Elsevier, North-Holland, Amsterdam, London, New York.

Frederiksen, A., 2004. Explaining individual job separations in a segregated labor market. Industrial Relations Section Working Paper No. 490, Princeton University.

Frederiksen, A., Honoré, B.E., Hu, L., 2006. Discrete time duration models with group-level heterogeneity. SIEPR Policy Paper No. 05-008 ⟨http://siepr.stanford.edu/papers/pdf/05-08.pdf⟩.

Heckman, J.J., 1981a. Heterogeneity and state dependence. In: Rosen, S. (Ed.), Studies in Labor Markets.

Heckman, J.J., 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Panel Data with Econometric Applications, pp. 179–195.

Heckman, J.J., 1981c. Statistical models for discrete panel data. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Panel Data with Econometric Applications. pp. 114–178.

Heckman, J.J., Honoré, B.E., 1990. The empirical content of the roy model. Econometrica 58, 1121–1149.

Holt, J., Prentice, R., 1974. Survival analysis in twin studies and matched pair experiments. Biometrika 61, 17–30.

Honoré, B.E., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68, 839–874.

Horowitz, J.L., 1992. A smoothed maximum score estimator for the binary response model. Econometrica 60, 505–531.

Hougaard, P., 2000. Analysis of Multivariate Survival Data. Springer, New York.

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

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

Light, A., Ureta, M., 1992. Panel estimates of male and female job turnover behaviour: can female nonquitters be identified? Journal of Labour Economics 10, 156–181.
Lynch, L.M., 1992. Differential effects of post-school training on early career mobility. Working Paper, National Bureau of Economic Research.
Manski, C.F., 1975. The maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3, 205–228.
Manski, C.F., 1987. Semiparametric analysis of random effects linear models from binary panel data. Econometrica 55, 357–362.
Meyer, B.D., 1990. Unemployment insurance and unemployment spells. Econometrica 58, 757–782.
Prentice, R.L., Gloeckler, L.A., 1978. Regression analysis of grouped survival data with application to breast cancer data. Biometrics 34, 57–67.
Rasch, G., 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Denmarks Pædagogiske Institut, Copenhagen.
Ridder, G., Tunali, I., 1999. Stratified partial likelihood estimation. Journal of Econometrics 92, 193–232.
Royalty, A., 1998. Job-to-job and job-to-nonemployment turnover by gender and education level. Journal of Labour Economics 14, 392–443.
Sastry, N., 1997. A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. Journal of the American Statistical Association 92, 426–435.
Thomas, A., 2002. Consistent estimation of discrete-choice models for panel with multiplicative effects. LEERNA-INRA, Université des Sciences Sociales de Toulouse.