



CAM

**Centre for Applied
Microeconometrics**

**Institute of Economics
University of Copenhagen**

<http://www.econ.ku.dk/CAM/>

**Estimation of Discrete Time Duration Models
with Grouped Data**

Bo E. Honoré
Luoja Hu

2004-07

**The activities of CAM are financed by a grant from
The Danish National Research Foundation**

Estimation of Discrete Time Duration Models with Grouped Data*

Bo E. Honoré[†] Luojia Hu[‡]

This Version: April 2004

Comments Welcome

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 child mortality allowing for family specific effects.

Keywords: Panel Data, Discrete Choice, Duration Models.

JEL Classification: C23, C33, C41.

1 Introduction

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. See

*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).

[†]Mailing Address: Department of Economics, Princeton University, Princeton, NJ 08544-1021. Email: honore@Princeton.edu.

[‡]Mailing Address: Department of Economics, Northwestern University, Evanston, IL 60208-2600. Email: luojiahu@Northwestern.edu.

for example Cox (1958), Heckman (1981a, 1981b, 1981c), Chamberlain (1985) or Honoré and Kyriazidou (2000). 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. This is in the spirit of classical duration models where the dynamics is captured through duration dependence (see Lancaster (1990)). The main contribution of the paper is to propose estimators that allow for group specific effect in parametric and semiparametric versions of the model.

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

$$y_{it} = 1 \{x'_{it}\beta + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it} \geq 0\} \quad (1)$$

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, (3) 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. See e.g., Heckman and Borjas (1980). 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} .

Here we consider a model in which an individual occupies a certain state and the outcome of interest is whether the individual leaves the state at time t . Notationally, $y_{\ell t} = 1$ will be used to describe the event that individual ℓ leaves the state at calendar time t , and we model this by

$$y_{\ell t} = 1 \{x'_{\ell t}\beta + \delta_{S_{\ell t}} + \alpha_{\ell} + \varepsilon_{\ell t} \geq 0\} \quad (2)$$

where $S_{\ell t}$ denotes the number of periods since the individual entered the state. The maintained assumption is that we observe a sample of individuals, ℓ , that is grouped in such a way that the individual-specific effect is the same across some ℓ . 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.

Throughout this paper, 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.

2 The Model and Estimator

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 $\Upsilon_{ji} = S_{ji1} + T_{ji}$.

As mentioned in the introduction, 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. Hence the model is

$$y_{jit} = 1 \{x'_{jit}\beta + \delta_{S_{jit}} + \alpha_i + \varepsilon_{jit} \geq 0\}, \quad t = 1, \dots, \bar{t}, \quad j = 1, \dots, J \quad i = 1, \dots, n \quad (3)$$

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 $\{y_{jit} : t = 1, \dots, \bar{t}, j = 1, \dots, J\}$ and $\{y_{jit} : t = 1, \dots, \bar{t}\}$, 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 $\{x'_{jit} : t = 1, \dots, T_{ji}, j = 1, \dots, J, i = 1, \dots, n\}$.

In what follows, we will assume that J is the same across groups. This can be easily relaxed provided that J is exogenous (formally, the assumptions below have to hold conditional on J).

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 $\left\{ \alpha_i, \{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leq t}, \{\varepsilon_{kis}\}_{s\leq t+\tau, k\neq j}, \{x_{kis}\}_{s\leq t+\tau, k\neq j}, \{S_{ki1}\}_{k=1}^J \right\}$ for some known τ .

This assumption corresponds to the logit assumption used in Rasch (1960), Cox (1958), Andersen (1970), Chamberlain (1985) and Honoré and Kyriazidou (2000).

The next assumption generalizes Assumption 2a in much the same way that Manski (1987) generalized Rasch's logit model with individual specific effects.

Assumption 2b. For some known τ ($\tau \geq 0$), and conditionally on z_i , $\{\varepsilon_{jit}\}_{j=1}^J$ are independent of each other and of $\left\{ \{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leq t}, \{\varepsilon_{kis}\}_{s\leq t+\tau, k\neq j}, \{x_{kis}\}_{s\leq t+\tau, k\neq j} \right\}$ for $t = 1, \dots, T$, 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} is allowed to vary across i .

It will also be relevant to consider a generalization of Assumption 2b that allows the distribution of ε_{jit} to depend on S_{jit} . This will make the duration dependence component of the model much less parametric.

Assumption 2c. For some known τ ($\tau \geq 0$), and conditionally on z_i , $\{\varepsilon_{jit}\}_{j=1}^J$ are independent of each other and of $\left\{ \{\varepsilon_{jis}\}_{s<t}, \{x_{jis}\}_{s\leq t}, \{\varepsilon_{kis}\}_{s\leq t+\tau, k\neq j}, \{x_{kis}\}_{s\leq t+\tau, k\neq j} \right\}$. Moreover, the distributions of ε_{jit} and ε_{lis} 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. See section 2.4

For a given individual, Assumptions 2a, 2b and 2c do 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 (3) differently from the other explanatory variables. However, the notation in (3) 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 specific¹.

For now assume that $J = 2$. We then have

Lemma 1 *Let t_1 and t_2 be arbitrary with $|t_1 - t_2| \leq \tau$. 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*

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

under Assumption 2b

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

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

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

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

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

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

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

subject to a scale normalization.

¹In this version of the paper, we assume that the calendar time for the first observation is the same for all individuals. This assumption is easily relaxed at the cost of slightly more cumbersome notation.

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

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

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 .

2.1 Group-Specific δ

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), (5) and (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), Honoré and D'Addio (2003). Those papers consider models with second order state dependence where the first order is allowed to be individual-specific.

2.2 Group-Specific x

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 .

2.3 Censoring

Covariate-dependent censoring is not a problem provided that it is independent of the ε 's. Specifically, let C_{ji} be the censoring time for individual j in group i . The argument above then applies if assumptions 2a, 2b and 2c are modified to

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

Assumption 2b'. For some known τ ($\tau \geq 0$), and conditionally on z_i , $\{\varepsilon_{jit}\}_{j=1}^J$ are independent of each other and of $\left\{ \{\varepsilon_{jis}\}_{s < t}, \{x_{jis}\}_{s \leq t}, \{\varepsilon_{kis}\}_{s \leq t+\tau, k \neq j}, \{x_{kis}\}_{s \leq t+\tau, k \neq j}, \{C_{ki}\}_{k=1}^J \right\}$ for $t = 1, \dots, T$, and the conditional distributions of $\{\varepsilon_{jit}\}_{j=1, t=1}^{J, T}$ are identical.

Assumption 2c'. For some known τ ($\tau \geq 0$), and conditionally on z_i , $\{\varepsilon_{jit}\}_{j=1}^J$ are independent of each other and of $\left\{ \{\varepsilon_{jis}\}_{s < t}, \{x_{jis}\}_{s \leq t}, \{\varepsilon_{kis}\}_{s \leq t+\tau, k \neq j}, \{x_{kis}\}_{s \leq t+\tau, k \neq j}, \{C_{ki}\}_{k=1}^J \right\}$. Moreover, the distributions of ε_{jit} and ε_{lis} 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

$$\begin{aligned} & \sum_{i=1}^n \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1 \left\{ |t_1 - t_2| \leq \tau, t_1 < \min_k \{C_{ki}\}, t_2 < \min_k \{C_{ki}\} \right\} \\ & \quad (1 \{T_{1i} = t_1, T_{2i} > t_2\} + 1 \{T_{1i} > t_1, T_{2i} = t_2\}) \\ & \quad \log \left(\frac{\exp((x_{1it_1} - x_{2it_2})\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))^{1\{T_{1i}=t_1, T_{2i}>t_2\}}}{1 + \exp((x_{1it_1} - x_{2it_2})\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}))} \right) \end{aligned}$$

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

$$\begin{aligned} & \sum_{i=1}^n \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1 \left\{ |t_1 - t_2| \leq \tau, t_1 < \min_k \{C_{ki}\}, t_2 < \min_k \{C_{ki}\} \right\} \\ & \quad \cdot 1 \{T_{1i} = t_1, T_{2i} > t_2\} \cdot 1 \{(x_{1it_1} - x_{2it_2})\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) > 0\} \\ & \quad + 1 \left\{ |t_1 - t_2| \leq \tau, t_1 < \min_k \{C_{ki}\}, t_2 < \min_k \{C_{ki}\} \right\} \\ & \quad \cdot 1 \{T_{1i} > t_1, T_{2i} = t_2\} \cdot 1 \{(x_{1it_1} - x_{2it_2})\beta + (\delta_{t_1+S_{1i1}} - \delta_{t_2+S_{2i1}}) < 0\} \end{aligned}$$

subject to a scale normalization.

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

$$\begin{aligned} & \sum_{i=1}^n \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1 \left\{ |t_1 - t_2| \leq \tau, t_1 + S_{1i1} = t_2 + S_{2i1}, t_1 < \min_k \{C_{ki}\}, t_2 < \min_k \{C_{ki}\} \right\} \\ & \cdot (1 \{T_{1i} = t_1, T_{2i} > t_2\} \cdot 1 \{(x_{1it_1} - x_{2it_2})'\beta > 0\} + 1 \{T_{1i} > t_1, T_{2i} = t_2\} \cdot 1 \{(x_{1it_1} - x_{2it_2})'\beta < 0\}) \end{aligned}$$

2.4 Pairwise Comparison Estimation When There Is No Group-Specific Effect

Following, for example Honoré and Powell (1994) it is natural to consider a non-panel version of the model in (3),

$$y_{it} = 1 \{x'_{it}\beta + \delta_{S_{it}} + \varepsilon_{it} \geq 0\}, \quad t = 1, \dots, \bar{t}, \quad i = 1, \dots, n \quad (7)$$

and then apply the approach discussed earlier to all pairs of observations i_1 and i_2 . In a semiparametric case, this would lead to an estimator defined by minimizing

$$\begin{aligned} & \sum_{i_1 < i_2}^n \sum_{t_1=1}^{\bar{t}} \sum_{t_2=1}^{\bar{t}} 1 \{T_{i_1} = t_1, T_{i_2} > t_2\} \cdot 1 \{(x_{i_1 t_1} - x_{i_2 t_2})' \beta + (\delta_{t_1} - \delta_{t_2}) > 0\} \\ & + 1 \{T_{i_1} > t_1, T_{i_2} = t_2\} \cdot 1 \{(x_{i_1 t_1} - x_{i_2 t_2})' \beta + (\delta_{t_1} - \delta_{t_2}) < 0\} \end{aligned} \quad (8)$$

In the case where $\bar{t} = 1$, (3) is a standard discrete choice model, and in that case the objective function in (8) becomes

$$\sum_{i_1 < i_2}^n 1 \{y_{i_1} > y_{i_2}\} \cdot 1 \{(x_{i_1} - x_{i_2})' \beta > 0\} + 1 \{y_{i_1} < y_{i_2}\} \cdot 1 \{(x_{i_1} - x_{i_2})' \beta < 0\}$$

which is the objective function for Han (1987)'s maximum rank correlation estimator.

It is also possible to link (7) to a general monotone index model of the form

$$G(T_i^*) = x_i' \beta + \varepsilon_i \quad (10)$$

where G is continuous and strictly increasing and a discretized version of T_i^* is observed. (10) implies that²

$$\begin{aligned} P(T_i^* > t | x_i) &= P(G(T_i^*) > G(t) | x_i) \\ &= P(x_i' \beta + \varepsilon_i > G(t) | x_i) \\ &= 1 - F(G(t) - x_i' \beta) \end{aligned}$$

where F is the CDF for ε_i . This gives

$$P(T_i^* > t + 1 | x_i, T_i^* > t) = \frac{1 - F(G(t + 1) - x_i' \beta)}{1 - F(G(t) - x_i' \beta)}.$$

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 $x_i' \beta$. This means that one can write the event $T_i^* > t + 1 | x_i, T_i^* > t$ in the form $1 \{x_i' \beta > \eta_{it}\}$ for some random variable η_{it} which is independent of x_i and has CDF $\frac{1 - F(G(t+1) - \cdot)}{1 - F(G(t) - \cdot)}$. This has the same structure as (7) with time-invariant explanatory variables combined with a non-panel version of Assumption 2c. In other

²Expressions of the form $P(T_i^* > t | x_i) = 1 - F(a_t - x_i' \beta)$ can also be obtained without the assumption that G is continuous and strictly increasing. The discussion here can therefore be generalized to more general monotone transformation models (at the cost of additional notation).

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. The estimator that results from exploiting this insight will share many of the rank estimators proposed in the literature such as Han (1987), Cavanagh and Sherman (1998), Abrevaya (1999), Chen (2002) and Khan and Tamer (2004). However, it does not appear that the estimator based on the approach taken here will be a special case of any of them, or vice versa.

2.5 J>2

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

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

For the semiparametric cases in assumptions 2b, we get

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

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

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

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

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

where

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

For the semiparametric cases in Assumptions 2b, we get

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

if and only if

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

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

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

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.

2.6 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 (3) 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 parameter 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. In that case the distribution of y_i given (x_i, α_i) is

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

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

2.7 Comparison to Continuous Case

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

$$\lambda\left(t \mid \{x_{is}\}_{s \leq t}\right) = \lambda(t) \exp(x'_{it}\beta)$$

(see Kalbfleisch and Prentice (1980)). Cox's estimator (Cox (1972), Cox (1975)) 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 i 'th 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 $t_1 + S_{1i1} = t_2 + S_{2i1}$.

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 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.

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

$$y_{jit}^1 = 1 \left\{ x'_{jit} \beta^1 + \delta_{S_{jit}^1} + \alpha_i^1 + \varepsilon_{jit} \geq 0 \right\}, \quad t = 1, \dots, \bar{t}, \quad j = 1, \dots, J \quad i = 1, \dots, n$$

$$y_{jit}^2 = 1 \left\{ x'_{jit} \beta^2 + \delta_{S_{jit}^2} + \alpha_i^2 + \varepsilon_{jit} \geq 0 \right\}, \quad t = T_{ji}^1 + 1, \dots, \bar{t}, \quad j = 1, \dots, J \quad i = 1, \dots, n$$

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 individual i_1 and i_2 to construct conditional statements like the ones in the previous section.

3.1.2 Comparing Second Spells

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

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

$$P\left(A|A \cup B, x_{1it_1^2}, x_{2it_2^2}, z_i\right) = \frac{\exp\left(\left(x'_{1it_1^2} - x'_{2it_2^2}\right)\beta^2 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^2 - t_2^1}^2\right)}{1 + \exp\left(\left(x'_{1it_1^2} - x'_{2it_2^2}\right)\beta^2 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^2 - t_2^1}^2\right)}; \quad (11)$$

Under Assumption 2b, we have

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

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

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

Since (11), (12) and (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 $B = \{T_{1i}^2 > t_1^2, T_{2i}^2 = t_2^2\}$.

3.1.3 Comparing First Spells to Second Spells (Assuming $\alpha_i^1 = \alpha_i^2 = \alpha_i$)

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

Under Assumption 2a, we have

$$P\left(A|A \cup B, x_{1it_1^2}, x_{2it_2^1}, z_i\right) = \frac{\exp\left(x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta_{t_1^2-t_1^1}^2 - \delta_{t_2^1+S_{2i1}}^1\right)}{1 + \exp\left(x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta_{t_1^2-t_1^1}^2 - \delta_{t_2^1+S_{2i1}}^1\right)}; \quad (14)$$

and under Assumption 2b

$$P\left(A|A \cup B, x_{1it_1}, x_{2it_2}, z_i\right) \begin{cases} > \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta_{t_1^2-t_1^1}^2 - \delta_{t_2^1+S_{2i1}}^1 > 0, \\ = \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta_{t_1^2-t_1^1}^2 - \delta_{t_2^1+S_{2i1}}^1 = 0, \\ < \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 + \delta_{t_1^2-t_1^1}^2 - \delta_{t_2^1+S_{2i1}}^1 < 0. \end{cases} \quad (15)$$

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

$$P\left(A|A \cup B, x_{1it}, x_{2it}, z_i\right) \begin{cases} > \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 > 0, \\ = \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 = 0, \\ < \frac{1}{2} & \text{if } x'_{1it_1^2}\beta^2 - x'_{2it_2^1}\beta^1 < 0. \end{cases} \quad (16)$$

Since (14), (15) and (16) 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 $B = \{T_{1i}^2 > t_1^2, T_{2i}^1 = t_2^1\}$.

4 An Empirical Application

In this section we illustrate the proposed estimation method by an analysis of child mortality allowing for family specific effects. It is well-established in the demographic literature that child mortality can be influenced by factors that are common to all siblings within a family such as genetics and parental competence. To investigate this issue, we use the same child survival data as Sastry (1997). The data come from the *Presquisa Nacional sobre Saúde Materno-Infantil e Planejamento Familiar – Brasil*, a household survey of Brazil that was conducted in 1986 as part of the Demographic and Health Survey program. Retrospective maternity histories were collected from 5,892 women age 15-44 who reported a total of 12,356 births. Following Sastry (1997), we restrict our analysis to the 2,946 singleton births that occurred within 10 years of the survey and in the northeast region. The geographic restriction is motivated by the fact that the northeast is a high-fertility and high-mortality region, which provides a better setting in which to study the unobserved family effect because it is the variation among siblings' survival that allows us to estimate the model. For the period 1976-86, the infant mortality rate was 143.7 and the child

mortality rate was 164.4 in the northeast, compared to 53.0 and 58.6, respectively, in the rest of Brazil. In addition, previous studies have found that the relationship between covariates and survival chances in this region is different from patterns found elsewhere in Brazil. (see e.g., Sastry(1995, 1997) for detail).

We consider a set of covariates that are typical in the previous demographic studies of child mortality, including the child's age and gender, birth order, birth-spacing and maternal age (mother's age at the birth of the child). Summary statistics for these variables are reported in Table 1.

The distribution of children by family is reported in Table 2. The 2,946 singleton births in the sample occurred to 1,051 families. There is a substantial amount of clustering of observations by family. Over 90% of the children belong to families that have two or more children in the sample. The average number of children per family is 2.8. The magnitude of the family specific effect in the model is mainly determined by the number of deaths per family since children in families with multiple deaths face higher mortality risks. Of the 430 deaths in the sample, nearly 60% come from the 9% families with two or more deaths. This suggests that it is important to control for unobserved family specific effects in the analysis of the effects of covariates on child mortality.

We estimate three model specifications and the results are presented in Table 3. Model 1 is a logit regression with no correction for family effects. Models 2 and 3 allow for unobserved family specific effects. The difference between the latter two is the way in which mortality rates are modeled. In Model 2, the mortality rates are defined over five (unequally spaced) age intervals; while Model 3 specifies the monthly mortality rates but restricts the duration coefficients to be the same within each of the five age intervals. The estimated relationship between covariates and mortality in general follows similar patterns across the three models. The mortality rate monotonically decreases with child age and is higher for boys. High parity births and short interbirth intervals are associated with higher mortality risks. However, compared to Model 1, the magnitude of the covariate coefficients is greatly reduced once family effects are controlled for in Models 2 and 3. Another significant difference is the estimated effect of maternal age. Without controlling for family effects, the risk ratio decreases steadily with maternal age in our sample of women between 14 and 55 years old. However, when family effects are accounted for in Models 2 and 3, the estimated child mortality risk ratio decreases with the mother's age initially but soon increases after age 30. In other words, children of women who give birth at a younger and older age experience higher mortality rates,

which is more consistent with previous studies.³

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.

We apply the methods developed here to an empirical analysis of child mortality with family specific effects. In future work, we also plan to consider other applications such as a study of employment durations with firm-specific effects.

References

- ABREVAYA, J. (1999): “Leapfrog Estimation of a Fixed-Effects Model with Unknown Transformation of the Dependent Variable,” *Journal of Econometrics*, 93, 203–228.
- ANDERSEN, E. B. (1970): “Asymptotic Properties of Conditional Maximum Likelihood Estimators,” *Journal of the Royal Statistical Society, Series B*, 32, 283–301.
- CAVANAGH, C. L., AND R. P. SHERMAN (1998): “Rank Estimators of Monotonic Index Models,” *Journal of Econometrics*, 84, 351–381.
- 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, no. 10

³See, for example, Sastry (1997). Sastry estimated a random effect proportional hazard model which allows for multi-level clustering by family and by community.

- in Econometric Society Monographs series,, pp. 3–38. Cambridge University Press, Cambridge, New York and Sydney.
- CHEN, S. (2002): “Rank Estimation of Transformation Models,” *Econometrica*, 70, 1683–1696.
- COX, D. R. (1958): “The Regression Analysis of Binary Sequences,” *Journal of the Royal Statistical Society*, 20, 215–242.
- (1972): “Regression models and life-tables,” *J. Royal Statistical Soc. Series B-statistical Methodology*, 34.
- (1975): “Partial likelihood,” *Biometrika*, 62.
- HAN, A. (1987): “Nonparametric Analysis of a Generalized Regression Model,” *Journal of Econometrics*, 35, 303–316.
- HECKMAN, J. J. (1981a): “Heterogeneity and State Dependence,” *Studies in Labor Markets*, S. Rosen (ed).
- (1981b): “The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process,” *Structural Analysis of Discrete Panel Data with Econometric Applications*, C. F. Manski and D. McFadden (eds), pp. 179–195.
- (1981c): “Statistical Models for Discrete Panel Data,” *Structural Analysis of Discrete Panel Data with Econometric Applications*, C. F. Manski and D. McFadden (eds), pp. 114–178.
- HECKMAN, J. J., AND G. J. BORJAS (1980): “Does Unemployment cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence,” *Economica*, 47, 247–283.
- HECKMAN, J. J., AND B. E. HONORÉ (1990): “The Empirical Content of The Roy Model,” *Econometrica*, 58, 1121–1149.
- HONORÉ, B. E., AND A. C. D’ADDIO (2003): “Duration Dependence and Time Varying Variables in Discrete Time Duration Models,” Princeton University.
- HONORÉ, B. E., AND E. KYRIAZIDOU (2000): “Panel Data Discrete Choice Models with Lagged Dependent Variables,” *Econometrica*, 68, 839–874.

- HONORÉ, B. E., AND J. L. POWELL (1994): “Pairwise Difference Estimators of Censored and Truncated Regression Models,” *Journal of Econometrics*, 64, 241–78.
- KALBFLEISCH, J. D., AND R. L. PRENTICE (1980): *The Statistical Analysis of Failure Time Data*. Wiley, New York.
- KHAN, S., AND E. TAMER (2004): “Partial Rank Estimation of Duration Models with General Forms of Censoring,” University of Rochester and Princeton University.
- LANCASTER, T. (1990): *The Econometric Analysis of Transition Data*. Cambridge University Press.
- MANSKI, C. (1987): “Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data,” *Econometrica*, 55, 357–62.
- MANSKI, C. F. (1975): “The Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 3, 205–228.
- RASCH, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests*. Denmark's Pædagogiske Institut, Copenhagen.
- RIDDER, G., AND I. TUNALI (1999): “Stratified Partial Likelihood Estimation,” *Journal of Econometrics*, 92, 193–232.
- SASTRY, N. (1995): “Community Characteristics, Individual Attributes and Child Survival in Brazil: A Multilevel Hazard Model Analysis,” Ph.D. dissertation, Princeton University.
- (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.

6 Appendix

6.1 Derivation of Lemma 1

Let t_1 and t_2 be arbitrary with $|t_1 - t_2| \leq \tau$, 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 $t_1 < t_2$.

$$\begin{aligned}
& P(A, \{x_{1it}\}_{t=2}^{t_1}, \{x_{2it}\}_{t=2}^{t_2} | z_i) \\
= & P_1(y_{1i1} = 0, y_{2i1} = 0 | z_i) \\
& \cdot p_2(x_{1i2}, x_{2i2} | z_i, y_{1i1} = 0, y_{2i1} = 0) \\
& \cdot \dots \\
& \cdot \dots \\
& \cdot P_{t_1} \left(y_{1it_1} = 1, y_{2it_1} = 0 \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1}, \{y_{1is} = 0, y_{2is} = 0\}_{s < t_1} \right) \\
& \cdot p_{t_1+1} \left(x_{2it_1+1} \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leq t_1} \right) \\
& \cdot P_{t_1+1} \left(y_{2it_1+1} = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_1+1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leq t_1} \right) \\
& \cdot p_{t_1+2} \left(x_{2it_1+2} \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_1+1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leq t_1+1} \right) \\
& \cdot P_{t_1+2} \left(y_{2it_1+2} = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_1+2}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leq t_1+1} \right) \\
& \cdot \dots \\
& \cdot \dots \\
& \cdot p_{t_2} \left(x_{2it_2} \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_2-1}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leq t_2-1} \right) \\
& \cdot P_{t_2} \left(y_{2it_2} = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_2}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leq t_2-1} \right),
\end{aligned}$$

and

$$\begin{aligned}
& P(B, \{x_{1it}\}_{t=2}^{t_1}, \{x_{2it}\}_{t=2}^{t_2} | z_i) \\
= & P_1(y_{1i1} = 0, y_{2i1} = 0 | z_i) \\
& \cdot p_2(x_{1i2}, x_{2i2} | z_i, y_{1i1} = 0, y_{2i1} = 0) \\
& \cdot \dots \\
& \cdot \dots \\
& \cdot P_{t_1} \left(y_{1it_1} = 0, y_{2it_1} = 0 \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1}, \{y_{1is} = 0, y_{2is} = 0\}_{s < t_1} \right) \\
& \cdot p_{t_1+1} \left(x_{2it_1+1} \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1}, \{y_{1is} = 0\}_{s \leq t_1}, \{y_{2is} = 0\}_{s \leq t_1} \right) \\
& \cdot P_{t_1+1} \left(y_{2it_1+1} = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_1+1}, \{y_{1is} = 0\}_{s \leq t_1}, \{y_{2is} = 0\}_{s \leq t_1} \right) \\
& \cdot p_{t_1+2} \left(x_{2it_1+2} \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_1+1}, \{y_{1is} = 0\}_{s \leq t_1}, \{y_{2is} = 0\}_{s \leq t_1+1} \right)
\end{aligned}$$

$$\begin{aligned}
& \cdot P_{t_1+2} \left(y_{2it_1+2} = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_1+2}, \{y_{1is} = 0\}_{s \leq t_1}, \{y_{2is} = 0\}_{s \leq t_1+1} \right) \\
& \dots \\
& \dots \\
& \cdot p_{t_2} \left(x_{2it_2} \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_2-1}, \{y_{1is} = 0\}_{s \leq t_1}, \{y_{2is} = 0\}_{s \leq t_2-1} \right) \\
& \cdot P_{t_2} \left(y_{2it_2} = 1 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_2}, \{y_{1is} = 0\}_{s \leq t_1}, \{y_{2is} = 0\}_{s \leq t_2-1} \right).
\end{aligned}$$

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

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

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

Either way one concludes that

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

where

$$\begin{aligned}
a_1 & = P_{t_1} \left(y_{1it_1} = 1, y_{2it_1} = 0 \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1}, \{y_{1is} = 0, y_{2is} = 0\}_{s < t_1} \right) \\
& \quad \cdot P_{t_2} \left(y_{2it_2} = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_2}, \{y_{1is} = 0\}_{s < t_1}, y_{1it_1} = 1, \{y_{2is} = 0\}_{s \leq t_2-1} \right) \\
a_2 & = P_{t_1} \left(y_{1it_1} = 0, y_{2i1} = 0 \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1}, \{y_{1is} = 0, y_{2is} = 0\}_{s < t_1} \right) \\
& \quad \cdot P_{t_1} \left(y_{2it_2} = 1 \mid z_i, \{x_{1is}\}_{s \leq t_1}, \{x_{2is}\}_{s \leq t_2}, \{y_{1is} = 0\}_{s \leq t_1}, \{y_{2is} = 0\}_{s \leq t_2-1} \right).
\end{aligned}$$

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

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

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 $t_1 + S_{1i1} = t_2 + S_{2i1}$

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

6.2 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.

6.2.1 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.

6.2.2 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 $|t_1^2 - t_2^1| \leq \tau$, 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)

$$\begin{aligned}
& P\left(A, \{x_{1it}\}_{t=2}^{t_1^1+t_1^2}, \{x_{2it}\}_{t=2}^{t_2^1} \mid z_i\right) \\
&= P_1\left(y_{1i1}^1 = 0, y_{2i1}^1 = 0 \mid z_i\right) \\
&\quad \cdot p_2\left(x_{1i2}, x_{2i2} \mid z_i, y_{1i1}^1 = 0, y_{2i1}^1 = 0\right) \\
&\quad \cdot \dots \\
&\quad \cdot \dots \\
&\quad \cdot P_{t_1^1}^1\left(y_{1it_1^1}^1 = 1, y_{2it_1^1}^1 = 0 \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1^1}, \{y_{1is}^1 = 0, y_{2is}^1 = 0\}_{s < t_1^1}\right) \\
&\quad \cdot p_{t_1^1+1}^1\left(x_{1it_1^1+1}, x_{2it_1^1+1} \mid z_i, \{x_{1is}, x_{2is}\}_{s \leq t_1^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{2is}^1 = 0\}_{s \leq t_1^1}\right) \\
&\quad \cdot P_{t_1^1+1}^2\left(y_{1it_1^1+1}^2 = 0, y_{2it_1^1+1}^1 = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1^1+1}, \{x_{2is}\}_{s \leq t_1^1+1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, \right. \\
&\quad \left. y_{1it_1^1}^1 = 1, \{y_{2is}^1 = 0\}_{s \leq t_1^1}\right) \\
&\quad \cdot p_{t_1^1+2}^1\left(x_{1it_1^1+2}, x_{2it_1^1+2} \mid z_i, \{x_{1is}\}_{s \leq t_1^1+1}, \{x_{2is}\}_{s \leq t_1^1+1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \right. \\
&\quad \left. y_{1it_1^1+1}^2 = 0, \{y_{2is}^1 = 0\}_{s \leq t_1^1+1}\right) \\
&\quad \cdot P_{t_1^1+2}^2\left(y_{1it_1^1+2}^2 = 0, y_{2it_1^1+2}^1 = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1^1+2}, \{x_{2is}\}_{s \leq t_1^1+2}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \right.
\end{aligned}$$

$$\begin{aligned}
& y_{1it_1+1}^2 = 0, \{y_{2is}^1 = 0\}_{s \leq t_1^1+1}) \\
& \dots \\
& \dots \\
& \cdot p_{t_2^1} \left(x_{1it_2^1}, x_{2it_2^1} \mid z_i, \{x_{1is}\}_{s \leq t_2^1-1}, \{x_{2is}\}_{s \leq t_2^1-1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_2^1-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1-1} \right) \\
& \cdot P_{t_2^1} \left(y_{1it_2^1}^2 = 0, y_{2it_2^1}^1 = 0 \mid z_i, \{x_{1is}\}_{s \leq t_2^1}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_2^1-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1-1} \right) \\
& \cdot p_{t_2^1+1} \left(x_{1it_2^1+1}, x_{2it_2^1+1} \mid z_i, \{x_{1is}\}_{s \leq t_2^1}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_2^1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1} \right) \\
& P_{t_2^1+1} \left(y_{1it_2^1+1}^2 = 0 \mid z_i, \{x_{1is}\}_{s \leq t_2^1+1}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_2^1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1} \right) \\
& \dots \\
& \dots \\
& \cdot p_{t_1^2} \left(x_{1it_1^2} \mid z_i, \{x_{1is}\}_{s \leq t_1^2-1}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_1^2-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1} \right) \\
& P_{t_1^2} \left(y_{1it_1^2}^2 = 1 \mid z_i, \{x_{1is}\}_{s \leq t_1^2}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_1^2-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1} \right)
\end{aligned}$$

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

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

where

$$\begin{aligned}
a_1 & = P_{t_1^2} \left(y_{1it_1^2}^2 = 0, y_{2it_1^2}^1 = 0 \mid z_i, \{x_{1is}\}_{s \leq t_2^1}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_1^2-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1-1} \right) \\
& \cdot P_{t_1^2} \left(y_{1it_1^2}^2 = 1 \mid z_i, \{x_{1is}\}_{s \leq t_1^2}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_1^2-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1-1} \right)
\end{aligned}$$

$$\begin{aligned}
& \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_1^2-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1} \\
&= \left(1 - F_{t_2^1} \left(x'_{1it_2^1} \beta^1 + \delta_{t_2^1 - t_1^1}^1 + \alpha_i^1\right)\right) \cdot \left(1 - F_{t_2^1} \left(x'_{2it_2^1} \beta^1 + \delta_{t_2^1 + S_{2i1}}^1 + \alpha_i^1\right)\right) \\
&\quad \cdot F_{t_1^2} \left(x'_{1it_1^2} \beta^2 + \delta_{t_1^2 - t_1^1}^2 + \alpha_i^2\right) \\
a_2 &= P_{t_2^1} \left(y_{1it_2^1}^2 = 0, y_{2it_2^1}^1 = 1 \mid z_i, \{x_{1is}\}_{s \leq t_2^1}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \right. \\
&\quad \left. \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_1^2-1}, \{y_{2is}^1 = 0\}_{s \leq t_2^1-1}\right) \\
&\quad \cdot P_{t_1^2} \left(y_{1it_1^2}^2 = 0 \mid z_i, \{x_{1is}\}_{s \leq t_1^2}, \{x_{2is}\}_{s \leq t_2^1}, \{y_{1is}^1 = 0\}_{s < t_1^1}, y_{1it_1^1}^1 = 1, \right. \\
&\quad \left. \{y_{1is}^2 = 0\}_{s=t_1^1+1}^{t_1^2-1}, \{y_{2is}^1 = 0\}_{s < t_2^1}, y_{2it_2^1}^1 = 1\right) \\
&= \left(1 - F_{t_2^1} \left(x'_{1it_2^1} \beta^1 + \delta_{t_2^1 - t_1^1}^1 + \alpha_i^1\right)\right) \cdot F_{t_2^1} \left(x'_{2it_2^1} \beta^1 + \delta_{t_2^1 + S_{2i1}}^1 + \alpha_i^1\right) \\
&\quad \cdot \left(1 - F_{t_1^2} \left(x'_{1it_1^2} \beta^2 + \delta_{t_1^2 - t_1^1}^2 + \alpha_i^2\right)\right)
\end{aligned}$$

so

$$P(A \mid A \cup B, \{x_{1it}\}_{t=1}^{t_1}, \{x_{2it}\}_{t=1}^{t_2}, z_i) \quad (17)$$

$$\begin{aligned}
& \frac{\left(1 - F_{t_2^1} \left(x'_{2it_2^1} \beta^1 + \delta_{t_2^1 + S_{2i1}}^1 + \alpha_i^1\right)\right) \cdot F_{t_1^2} \left(x'_{1it_1^2} \beta^2 + \delta_{t_1^2 - t_1^1}^2 + \alpha_i^2\right)}{F_{t_2^1} \left(x'_{2it_2^1} \beta^1 + \delta_{t_2^1 + S_{2i1}}^1 + \alpha_i^1\right) \cdot \left(1 - F_{t_2^1} \left(x'_{1it_2^1} \beta^2 + \delta_{t_2^1 - t_1^1}^2 + \alpha_i^2\right)\right)} \\
&= \frac{\left(1 - F_{t_2^1} \left(x'_{2it_2^1} \beta^1 + \delta_{t_2^1 + S_{2i1}}^1 + \alpha_i^1\right)\right) \cdot F_{t_1^2} \left(x'_{1it_1^2} \beta^2 + \delta_{t_1^2 - t_1^1}^2 + \alpha_i^2\right)}{1 + \frac{F_{t_1^2} \left(x'_{2it_2^1} \beta^1 + \delta_{t_2^1 + S_{2i1}}^1 + \alpha_i^1\right) \cdot \left(1 - F_{t_2^1} \left(x'_{1it_2^1} \beta^2 + \delta_{t_2^1 - t_1^1}^2 + \alpha_i^2\right)\right)}{F_{t_2^1} \left(x'_{2it_2^1} \beta^1 + \delta_{t_2^1 + S_{2i1}}^1 + \alpha_i^1\right) \cdot \left(1 - F_{t_1^2} \left(x'_{1it_1^2} \beta^2 + \delta_{t_1^2 - t_1^1}^2 + \alpha_i^2\right)\right)}}
\end{aligned}$$

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 (17) implies that under Assumption 2a,⁴

$$P\left(A \mid A \cup B, x_{1it_1^2}, x_{2it_2^1}, z_i\right) = \frac{\exp\left(x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^1 + S_{2i1}}^1\right)}{1 + \exp\left(x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^1 + S_{2i1}}^1\right)}; \quad (18)$$

⁴In 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.

and under Assumption 2b

$$P(A|A \cup B, x_{1it_1}, x_{2it_2}, z_i) \begin{cases} > \frac{1}{2} & \text{if } x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^1 + S_{2i1}}^1 > 0, \\ = \frac{1}{2} & \text{if } x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^1 + S_{2i1}}^1 = 0, \\ < \frac{1}{2} & \text{if } x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^1 + S_{2i1}}^1 < 0. \end{cases} \quad (19)$$

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

$$P(A|A \cup B, x_{1it}, x_{2it}, z_i) \begin{cases} > \frac{1}{2} & \text{if } x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 > 0 \\ = \frac{1}{2} & \text{if } x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 = 0 \\ < \frac{1}{2} & \text{if } x'_{1it_1^2} \beta^2 - x'_{2it_2^1} \beta^1 < 0 \end{cases} . \quad (20)$$

Since (18), (19) and (20) do not depend on t_1^1 and t_2^1 , the same statements are true if we redefine A and B as $A = \{T_{1i}^2 = t_1^2, T_{2i}^1 > t_2^1\}$ and $B = \{T_{1i}^2 > t_1^2, T_{2i}^1 = t_2^1\}$.

6.2.3 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 $|t_1^2 - t_2^2| \leq \tau$, and recall that z_i denotes the set of predetermined variables for group i at the beginning of the sample.

Consider the two events $A = \{T_{1i}^1 = t_1^1, T_{1i}^2 = t_1^2, T_{2i}^1 = t_2^1, T_{2i}^2 > t_2^2\}$ and $B = \{T_{1i}^1 = t_1^1, T_{1i}^2 > t_1^2, T_{2i}^1 = t_2^1, T_{2i}^2 = t_2^2\}$. Mimicking the calculations above we find that under Assumption 2a,

$$P(A|A \cup B, x_{1it_1^2}, x_{2it_2^2}, z_i) = \frac{\exp\left(\left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^2 - t_2^1}^2\right)}{1 + \exp\left(\left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^2 - t_2^1}^2\right)}; \quad (21)$$

and under Assumption 2b

$$P(A|A \cup B, x_{1it_1^2}, x_{2it_2^2}, z_i) \begin{cases} > \frac{1}{2} & \text{if } \left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^2 - t_2^1}^2 > 0, \\ = \frac{1}{2} & \text{if } \left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^2 - t_2^1}^2 = 0, \\ < \frac{1}{2} & \text{if } \left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 + \delta_{t_1^2 - t_1^1}^2 - \delta_{t_2^2 - t_2^1}^2 < 0. \end{cases} \quad (22)$$

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

$$P(A|A \cup B, x_{1it_1^2}, x_{2it_2^2}, z_i) \begin{cases} > \frac{1}{2} & \text{if } \left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 > 0, \\ = \frac{1}{2} & \text{if } \left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 = 0, \\ < \frac{1}{2} & \text{if } \left(x'_{1it_1^2} - x'_{2it_2^2}\right) \beta^2 < 0. \end{cases} \quad (23)$$

Since (21), (22) and (23) do not depend on t_1^1 , the same statements are true if we redefine A and B as $A = \{T_{1i}^2 = t_1^2, T_{2i}^2 > t_2^2\}$ and $B = \{T_{1i}^2 > t_1^2, T_{2i}^2 = t_2^2\}$.

Table 1. Summary Statistics: Births in Northeast Brazil 1976-1986

Variables	Mean or Percent in Category
Gender	
Male	50.58%
Maternal age (years)	
Mean	26.02
Standard deviation	6.01
Birth order/preceding birth interval ^a	
First birth	21.11%
Order 2-4/short	9.40%
Order 2-4/medium	20.20%
Order 2-4/long	13.27%
Order 5+/short	9.00%
Order 5+/medium	16.84%
Order 5+/long	10.18%
Number of births	2,946
Number of deaths	430

^a Previous birth interval length: short, <15 months; medium, 15-29 months; long, > 30 months.

Table 2. Distribution of Children by Family

Children per family	Death per family							Total
	0	1	2	3	4	5	6	
1	255	12						267
2	239	44	2					285
3	143	41	15	3				202
4	69	30	9	2	0			110
5	43	34	15	9	3	0		104
6	15	18	8	5	3	0	1	50
7	4	4	7	4	2	0	0	21
8	1	2	4	3	1	1	0	12
Total	769	185	60	26	9	1	1	1,051

Table 3. Estimates of Child Mortality Model With and Without Family Effects

Variables	Model 1		Model 2		Model 3	
Age (months) ^a						
1-5	-2.649	(0.422)	0.462	(0.141)	-1.215	(0.149)
6-11	-3.533	(0.429)	0.009	(0.160)	-1.824	(0.167)
12-23	-5.229	(0.457)	-1.076	(0.227)	-3.533	(0.237)
24+	-7.901	(0.469)	-1.200	(0.255)	-5.069	(0.255)
Birth ordering /preceding birth interval ^b						
First	-0.897	(0.278)	-0.264	(0.267)	-0.589	(0.324)
Order 2-4/short	0.865	(0.294)	0.009	(0.225)	-0.070	(0.289)
Order 2-4/long	-1.237	(0.322)	-0.550	(0.292)	-0.811	(0.321)
Order 5+/short	1.206	(0.321)	0.172	(0.265)	-0.088	(0.350)
Order 5+/medium	0.591	(0.276)	0.129	(0.268)	-0.010	(0.376)
Order 5+/long	-0.412	(0.376)	-0.547	(0.387)	-0.788	(0.567)
Gender						
Male	0.229	(0.162)	0.178	(0.133)	0.106	(0.183)
Maternal age						
Linear effect	0.030	(0.108)	-0.447	(0.167)	-0.291	(0.193)
Squared effect	-0.0013	(0.0019)	0.007	(0.003)	0.005	(0.003)

^a Omitted category: age < 1 month.

^b Omitted category: order 2-4/medium.