Social Interaction Effects in Duration Models

Bettina Drepper,* Georgios Effraimidis†

January 2013

(Job Market Paper)

Abstract

We introduce a new strategy for the identification of social interaction effects from grouped transition data exploiting information on the timing of transitions, with applications including starting to smoke/use drugs for the first time within a group of peers. In our approach, we jointly model the transition hazards of all members of a peer group, allowing for two sources of dependence: (1) Once a group member transitions, this directly affects the subsequent transition hazard of their peers (social interaction effect). Such effects may differ across group members, covariates and over successive transitions in the group; (2) Group members may have similar unobserved characteristics (correlated effect). This duration framework allows overcoming the reflection problem (Manski, 1993) in the presence of correlated effects, without making use of an exclusion restriction or instrument. An identification result of our model is presented, constituting an extension of the timing of events approach. We apply our model to the first-time use of marijuana among siblings growing up together in American households, using data from the NLSY79 and find that first-time drug use by the oldest sibling has a significant positive effect on the subsequent drug use behavior of the younger siblings.

*University of Mannheim, Center for Empirical Economics and Econometrics. Address for correspondence: L7, 3-5, 68161 Mannheim, Germany. Email: bdrepper@mail.uni-mannheim.de

†University of Southern Denmark Institut for Virksomhedsledelse og Økonomi. Address for correspondence: Campusvej 55, 5230 Odense M, Denmark. Email: georgiosnl@gmail.com

Keywords: Social interactions, peer effects, reflection problem, multivariate duration analysis, hazard rate, timing of events

Acknowledgements: We thank Gerard J. van den Berg, Petyo Bonev, Andrew Chesher, Bo Honoré, Enno Mammen, Áureo de Paula, Steffen Reinhold, Yuya Takahashi, Arne Uhlendorff and Andrea Weber for helpful comments. We also thank participants of seminars and conferences at Mannheim University, University College London, ESEM 2012 and the Joint Workshop of Empirical Economists and Sociologists in Mannheim for useful comments.

1 Introduction

The study of social interactions has been of constant interest in health and labor economics over the past two decades, with the main difficulty in the identification of social interactions laid out in the seminal work by Manski (1993). Labeled the reflection problem, in a reduced form linear model in which the reference group's average outcome measures the behavior of peers, it is difficult to determine if a person's behavior affects their peers or vice versa.[1] In this paper, we introduce a new strategy for the identification of social interaction effects from grouped transition data, using a multivariate duration framework. We jointly model the transition hazards of all members of a peer group, allowing for two sources of dependence conditional on observable characteristics: (1) Once a group member transitions, this may directly affect the subsequent transition hazards of the other group members ('social interaction effect'); (2) Group members may have similar unobserved characteristics ('correlated effect'). This definition of social interactions in terms of a lagged[2] effect in time allows overcoming the reflection problem in the presence of correlated effects without making use of an instrument (see Case and Katz, 1991; Monstad et al., 2011) or exclusion restriction, as suggested by Moffitt (2001). Furthermore, our approach allows studying social interactions in natural peer groups such as a circle of friends, work colleagues or neighborhoods, which are often the result of a self-selection process based on similar observable and unobservable characteristics. Additionally, social interaction effects are highly flexible within our model, and may differ across different group members, covariates and over successive transitions in the group.

In many applications of social interactions, the behavior of interest is characterized by a transition at a particular point in time following some entry point. Examples include the time at which a person purchases a new product following its release, or the age at which a person first has sexual intercourse, moves out of the neighborhood or starts/stops using drugs. In our empirical application, we study social interaction effects in the use of marijuana/hashish by siblings growing up together in American households[3]. Substance use is considered a highly

[1] Different versions of this model have been widely used in applications estimating peer effects (see Gaviria and Raphael, 2001; Sacerdote, 2001; Cohen-Cole and Fletcher, 2008; Brock and Durlauf, 2000).

[2] As we use a continuous time framework, the period between transition and response is practically zero in our identification result.

[3] In our application, instead of classical peer groups, we study teenagers growing up in the same household. Here, 'correlated effects' do not arise due to selection effects but rather to similar genetic factors and childhood

social behavior (see Gaviria and Raphael, 2001; Kawaguchi, 2003). When a teenager uses drugs for the first time, this could directly affect the subsequent behavior of their siblings in several different ways. This transition may cause their siblings to copy this behavior. Alternatively, a transition may reduce the stigma attached to the use of drugs, or simply raise curiosity. Besides these classical channels of social interaction or peer effects, a response could also be triggered by an information effect or the accessibility of drugs. In particular, the first transition within a group often constitutes a release of new information, and additionally, in the case of drug use, an effect of accessibility. As our approach allows distinguishing[4] between the effects of the first transition and subsequent transitions within a group, we can to some extend detect the existence of accessibility and information effects versus classical peer effects.[5]

In the model by Manski (1993), social interaction effects are assumed to be homogeneous across group members, i.e. the action of every group member has the same effect on any other member. In this paper, we show that the joined observation of transition times within a group allows identifying additional dynamics in a group of socially interacting individuals. Firstly, the degree to which a transition of a group member j affects the behavior of another member k may depend on the degree of social status/reputation of both members j and k within the group, as well as the combination of their observable characteristics x_j and x_k. For example, the oldest sibling may have a unique social role within the household, increasing the degree to which his behavior affects the younger siblings. At the same time, the oldest sibling may not be as strongly influenced by the behavior of his younger siblings. In our application, we find evidence for a significant influence of the behavior of the oldest sibling, but no evidence for an effect of a transition by a younger sibling.[6] Similarly, peers may more strongly affect each other if they have the same gender or belong to the same age group. Secondly, the strength of an effect may strongly depend on how many transitions have been experienced within the group up to this point. Since social interactions may exhibit different degrees of

---

effects.

[4]In our empirical analysis we have not implemented this distinction at this point.

[5]Although we are aware that our definition of 'social interaction effects' in this paper does not only capture classical peer effects/social interaction effects, we will use this terminology throughout this paper.

[6]A different application constitutes a supervisor who has a unique social role at their workplace, increasing the degree to which their behavior may affect his employees. At the same time, the behavior of employees may strongly influence other co-workers but not necessarily the supervisor.

contagiousness, we allow for the strength of the effect to increase/decrease or follow any other pattern with each additional transition experienced within the group.[7] This also captures the extreme case when no transitions have any effect, apart from the first. In our example of first time drug use, this pattern could arise if interaction effects are purely driven by the effect of new information or accessibility.

The identification of such patterns facilitates a deeper understanding of how social interaction effects evolve over time, depending on the composition of the peer group. It enables policy makers to intervene more effectively by targeting the key members of groups. If we consider a policy aimed at preventing the early drug use of teenagers, our model can be used to predict how drug use spreads throughout the group over time and how this pattern depends on the group member whose behavior is initially altered by such a policy.

Individuals often enter a (peer) group at different calendar dates. For example, new co-workers are hired, teenagers join a circle of friends/social network or new children are born in a household. The key members are often those who enter the group first, such as the oldest sibling being the first child in the household. In our main model specification, group members are labeled according to their order of entry.[8] Varying entry points play a crucial role in many applications, because they determine the different starting points of an underlying risk process faced by all individuals in the sample. In the case of siblings, this process represents the dependence of the risk to start using drugs on age. We also consider the case of a common entry point for all group members. One such example is the release of a new product, whereby after the day of release, all members of a peer group simultaneously start to face a certain risk of purchasing the new product.

In this paper, we present a multivariate mixed (proportional)[9] hazard type model that uses the information in the timing of transitions to identify social interaction effects in the presence of correlated unobserved characteristics. The idea of exploiting the timing of events

[7] Our model also includes the possibility of a negative interaction effect i.e. a transition of a group member decreases the probability of subsequent transitions within the group.

[8] This restricts the variation in entry dates to a setting with a predefined entry order, which complicates identification. In our main model specification, we focus on this case of ordered entry dates. The case of unrestricted variation in entry times is also discussed briefly. Our results can be extended to this case in a straightforward manner.

[9] In Section 2.3, we discuss conditions under which the proportionality assumption can be dropped, leading to a multivariate mixed hazard model.

to disentangle a causal effect from a selection effect is introduced by Abbring and Van den Berg (2003b), in the context of the evaluation of labor market programs. An extension to two full spells[10] is used in Van den Berg and Drepper (2011) in studying bereavement effects within twin pairs. In the special case of a group with two spells sharing a common entry date, our model reduces to this setting. An extension to multiple parallel spells raises several new issues that are not encountered in this two-spell setting, such as differences of interaction effects across different combinations of group members and how effects may change over subsequent transitions within the group. Furthermore, we account for different entry dates across members and discuss the relaxation of the proportionality assumption. In the following section, we present our identification results for this extended model.

There is a straightforward intuition for the identification of models exploiting the timing of events.[11] The process of successive transitions and responses of the transition hazards within a group generates distinct patterns in the data, which provides information on the existence of 'interaction effects' vs. 'correlated effects'. For instance, if transitions are observed within increasingly shorter intervals, irrespective of when the first transition occurs, such epidemics-type clustering of transitions indicates that the transitions of peers positively affect the subsequent transition hazard of the other group members (positive interaction effect). On the other hand, 'correlated effects' create heterogeneity across groups in the data. [12]

In the field of discrete choice models, social interaction effects are frequently captured by a penalty term for deviating from the behavior of other group members in the utility function (for an overview, see Blume et al., 2010; Brock and Durlauf, 2001). Honoré and De Paula (2010) introduce a model of two durations with an endogenous effect, building on a two player simultaneous game where the exit of one player increases the potential payoff of the other once they also exit. In contrast to this strand in the literature using equilibrium models with interdependent utility functions, we do not specify the underlying behavioral model of social interactions. Rather than assuming that individuals simultaneously decide to play best

[10]Abbring and Van den Berg (2003b) highlight that their model can be straightforwardly extended to a setting with two full spells, whereby the exit of each spell can affect the survival of the other (see also Abbring and Heckman, 2007; Freund, 1961).

[11]Abbring and Van den Berg (2003b) provide a similar intuition for the two spell setting.

[12]Here, in order to disentangle 'interaction-' from 'correlated effects', a crucial identifying assumption is that correlated unobservable characteristics remain constant over time.

responses, we understand social interactions as a dynamic process of successive actions and reactions within a group. A key feature of our approach is that the transition hazard of a group member may directly react in response to transitions of other members.[13]

In order to define a social interaction effect in terms of a response in the transition hazard, we assume that this response does not take place before the transition causing it has occurred ('no-anticipation' assumption, see Abbring and Van den Berg 2003b). This implies that individuals should not anticipate the action of fellow peers, or at least should not react to it before it is experienced. In applications where forward looking and strategic incentives dominate the behavior of group members, equilibrium models are more suitable for capturing such dynamics (see for example Honoré and De Paula, 2010). In contrast, our approach focuses on applications where a transition of a group member is comparable to an unanticipated shock that causes a systematic change in the behavior of the other members. We argue that the first-time substance use among siblings constitutes such an event. Teenagers are often influenced by sources outside the own household that are difficult to foresee by other household members. If a teenager is exposed to drugs at his school, the change in his behavior may subsequently affect his siblings at home.

We use data from the National Longitudinal Surveys (NLSY79) in our application, observing the first-time use of marijuana by 8,684 siblings in 5,810 American households, including 1,549 two-sibling households and 669 households with more than two siblings growing up together. We find that the first-time use of marijuana by the oldest sibling in the household has a significant positive effect on the subsequent drug use behavior of his younger siblings. However, we do not find evidence for an effect of a transition of a younger sibling. Females are more strongly influenced by the drug use behavior of their siblings than males.

In the next section we introduce our model of social interaction effects and present our identification results. In Section 3 we discuss the data set, estimation method and results of our application. We conclude in Section 4.

[13] This is a fundamental difference from Honoré and De Paula (2010), where this type of direct response in the hazard rate is ruled out.

2 A multiple-spell duration model with social interaction effects

In the following we introduce a model of three parallel spells ($J = 3$). We restrict attention to this three-spell case in this section, since all interesting dynamics occur within this setting. The extension to more than three spells is straightforward and will not be further discussed.

2.1 General framework

The three group members $j = 1, 2, 3$ enter into the origin state at member specific entry dates d_j. In our empirical example of first-time drug use, d_j denotes the calendar date at which sibling j reaches the threshold age after which he will be exposed to the risk of using drugs. To have a compact notation, we introduce the vector $d = (d_1\ d_2\ d_3)'$. Next, we denote by T_j the duration of member j until he transitions to the new state (e.g. the state of having used drugs). Furthermore, we introduce the μ-dimesnional vector $x \in \mathbb{X} \subseteq \mathbb{R}^\mu$ which holds all relevant observed covariates, member- and group-specific, that affect the realization of the duration variables. Additionally, the behavior of all group members is affected by unobservable influences denoted by the random vector $V = (V_1\ V_2\ V_3)'$ drawn from the non-degenerate trivariate cumulative density function G which does not depend on x and has support $\mathbb{V} \subseteq \mathbb{R}^3_+$.

We define our model in terms of conditional transition hazards of each duration T_j given the realization of the other two durations T_k, T_l, entry dates d, observable influences x and unobservable influences V_j

$$\theta_j(t|T_k, T_l, d, x, V_j) = \begin{cases} \lambda_{j,0}(t, d, x)V_j & \text{if } t \leq \min\{T_{jk}, T_{jl}\}, \\ \lambda_{j,k}(t|T_{jk}, d, x)V_j & \text{if } T_{jk} < t \leq T_{jl}, \\ \lambda_{j,l}(t|T_{jl}, d, x)V_j & \text{if } T_{jl} < t \leq T_{jk}, \\ \lambda_{j,kl}(t|T_{jk}, T_{jl}, d, x)V_j & \text{if } \max\{T_{jk}, T_{jl}\} < t. \end{cases} \tag{1}$$

with $T_{jk} := T_k + d_k - d_j$ for $j, k, l = 1, 2, 3$ such that $k \neq j \neq l \neq k$ and $k < l$.

The stochastic variable T_{jk} denotes the elapsed time between the entry of member j into the risk process and the transition of member k into the state of interest. In particular, if its value is negative (positive) then the transition of member k takes place before (after) the entry of member j.

The above model suggests a straightforward definition of the interaction effect functions as ratios of the conditional hazard rates in (1)

$$\delta_{j,k}(t|T_{jk},d,x) := \frac{\lambda_{j,k}(t|T_{jk},d,x)}{\lambda_{j,0}(t,d,x)} \tag{2}$$

$$\delta_{j,kl}(t|T_{jk},T_{jl},d,x) := \frac{\lambda_{j,kl}(t|T_{jk},T_{jl},d,x)}{\lambda_{j,q}(t|T_{jq},d,x)} \quad \text{for } q = \arg\min_{k,l}\{T_{jk},T_{jl}\} \tag{3}$$

with (2) representing the effect of the exit of member k on the hazard of member j and (3) the additional effect of the second exit on the hazard of member j. Note that, since the interaction effect functions are defined in terms of hazard rates conditional on the realization of V_j, they have a causal interpretation. The unobservable terms V_j drop in the ratios in (2) and (3). The functions $\lambda_{j,k}$ and $\lambda_{j,kl}$ are not directly observable from the data, since they are components of the conditional hazard rates $\theta_j(t|T_k,T_l,d,x,V_j)$. This poses an identification problem for the interaction or the functions $\delta_{j,k}$ and $\delta_{j,kl}$ which we will address in this section.

The identification results in this section build on the assumptions implied by the structure of model (1). Firstly, the unobservable influences $(V_1\ V_2\ V_3)'$ that are a source of the dependence between the three durations are assumed to be time-constant and enter the hazard rate multiplicatively, reflecting a reinforcing effect between observable and unobservable influences. The resulting mixed hazard structure is a popular choice in duration models . Secondly, in model (1) the effect of a transition of a member k enters the hazard rate of member j only after it occurs (for all $t > T_{jk}$). This assumption restricts the dependence structure between the three transitions T_1, T_2 and T_3 and is known as the 'no-anticipation' assumption (see Abbring and Van den Berg, 2003b). It plays a crucial role for the identification and estimation of model (2) as it allows to express the joint distribution of $\{T_1,T_2,T_3\}|\{d,x,V\}$ in terms of conditional distributions $\{T_j\}|\{T_k,T_l,d,x,V_j\}$. This allows to indirectly specify the joint distribution $\{T_1,T_2,T_3\}|\{d,x\}$ by specifying the conditional hazard rates in (1).

In this section, we discuss different sets of assumptions under which the interaction effect functions (2) and (3) in model (1) can be identified. We first consider the case of proportionality of the covariate effects leading to the popular mixed proportional hazard specification.

Model A: *Transition hazard of member j given T_k, T_l, d, x and V_j*

$$\theta_j(t|T_k, T_l, d, x, V_j) = \lambda_j(t)\phi_j(x)\delta_j(t|T_k, T_l, d, x)V_j$$

with social interaction effect functions

$$\delta_j(t|T_k, T_l, d, x) := \delta_{j,k}(t|T_{jk}, \mathcal{N}_k, x)^{I_{j,k}(t)}\delta_{j,l}(t|T_{jl}, \mathcal{N}_l, x)^{I_{j,l}(t)}\delta_{j,kl}(t|T_{jk}, T_{jl}, \mathcal{N}_{kl}, x)^{I_{j,kl}(t)},$$

where $\mathcal{N}_j := \sum_{s=1}^{3} \mathbb{I}(d_j + T_j > d_s)$, $\mathcal{N}_{kl} := \sum_{s=1}^{3} \mathbb{I}(d_q + T_q > d_s)$ *with* $q = arg\ max_{k,l}\{T_{jk}, T_{jl}\}$, $I_{j,k}(t) := \mathbb{I}(T_{jl} \geq t > T_{jk})$, $I_{j,kl}(t) := \mathbb{I}(max\{T_{jk}, T_{jll}\} < t)$ *with* $j, k, l = 1, 2, 3$ *such that* $k \neq j \neq l \neq k$ *and* $k < l$.

Here, $\mathbb{I}(.)$ is the indicator function. The variables $\mathcal{N}_k$ and $\mathcal{N}_{kl}$ are used to capture the size of the group at the calendar dates $d_k + T_k$ and $\max\{d_k + T_k, d_l + T_l\}$, respectively. Namely, the above specification allows the interaction effects to depend on the time of occurrence of the corresponding transition. In particular, $\mathcal{N}_j$ specifies the number of members who have entered the risk process at calendar date $d_j + T_j$ at which member j transitions. Similarly, $\mathcal{N}_{kl}$ gives the number of the members who have entered the risk process at the calendar date $\max\{d_k + T_k, d_l + T_l\}$, that is, when the second transition of member k or l occurs.

Before the first transition takes place, the hazard rates of the three durations are of the mixed proportional form. The function $\lambda_j(t)$ captures the duration dependence and $\phi_j(x)$ reflects the influence of observable member- and group-specific characteristics.

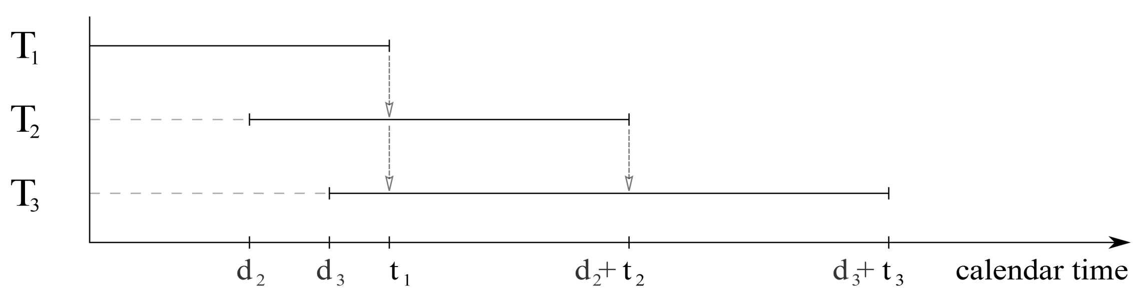

Figure 1: Example of three parallel spells in Model A: The first transition $T_1 = t_1$ occurs after the other two members have entered the risk process ($d_1 = 0 < d_2 < d_3 < t_1$). Member 2 is the second to transition at age t_2 and member 3 transitions last at age t_3.

To provide some intuition for model (1), we consider a concrete example in Figure 1. Here, the individual who is labeled as 1 (i.e. individual who enters the risk process first) transitions into the state of interest (e.g. use of drugs) first at calendar date t_1 ($T_1 < \min\{d_2+T_2, d_3+T_3\}$, with $T_1 = t_1$). By then, individuals labeled as 2 and 3 have both passed their threshold calendar date (d_2 and d_3, respectively, with $d_2 < d_3 < t_1$) and are at risk of transitioning into the state of interest.

Before the first transition has taken place at calendar date t_1, the transition hazard of member j is given by $\lambda_j(t)\phi_j(x)V_j$, for $j = 1, 2, 3$. After the first transition at calendar time t_1, the interaction effect functions $\delta_{2,1}(t|t_1-d_2, 3, x)$ and $\delta_{3,1}(t|t_1-d_3, 3, x)$ appear in the hazard rates of the two remaining durations T_2 and T_3 for all $t > t_1 - d_j$ for $j = 2, 3$ respectively. Next, we look at the effect of the second transition. Specifically, we have $T_2 = t_2$ with $t_2+d_2 > t_1 > d_3$. In this case, an additional interaction effect term $\delta_{3,12}(t|t_1 - d_3, t_2 + d_2 - d_3, 3, x)$ appears in the hazard of the surviving duration T_3 for all $t > t_2 + d_2 - d_3$. The interaction effect functions $\delta_{j,k}$ and $\delta_{j,kl}$ reflect that the transition of a group member affects the behavior of his fellow peers resulting in a potential change in their subsequent transition hazards.

To identify Model A, we will employ a set of certain assumptions that we formalize below.

Assumption A.1 *The function* $\phi_j : \mathbb{X} \to (0, \infty)$ *is such that it attains all values in an open connected subset of* $(0, \infty)$ *and also* $\phi_j(x^*) = 1$ *for some* $x^* \in \mathbb{X}$, *and* $j = 1, 2, 3$.

Assumption A.2 *The function* $\lambda_j : \mathbb{R}_+ \to (0, \infty)$ *is measurable and the integrated baseline hazard rate* $\Lambda_j(t) := \int_0^t \lambda_j(\omega)d\omega$ *exists and is finite for all* $t > 0$ *with* $\Lambda_j(t^*) = 1$ *for some particular* $t^* > 0$, $j = 1, 2, 3$.

Assumption A.3 *The* G *is does not depend on* x *and* d. *Moroever, for* $j = 1, 2, 3$, $\mathbb{E}(V_j) < \infty$

.

Assumption A.4 *For* $j, k, l = 1, 2, 3$ *such that* $k \neq j \neq l$ *and* $k < l$. *Let* $q = \arg\min_{k,l}\{T_{jk}, T_{jl}\}$ *and* $\pi(s, y) = \max\{0, \min\{s, y\}\}$. *The functions* $\delta_{j,k} : \mathbb{R}_+ \times \mathbb{R} \times \{1, 2, 3\} \times \mathbb{X} \to (0, \infty)$, *and*

$\delta_{j,kl} : \mathbb{R}_+ \times \mathbb{R}^2 \times \{1,2,3\} \times \mathbb{X} \rightarrow (0,\infty)$ *are measurable, ii) the quantities*

$$\Upsilon_{j,k}(t|s,\mathcal{N}_k,x) := \int_{\max\{0,s\}}^{t} \lambda_j(\omega)\delta_{j,k}(\omega|s,\mathcal{N}_k,x)d\omega,$$

$$\Delta_{j,k}(t|s,\mathcal{N}_k,x) := \int_0^t \delta_{j,k}(\omega|s,\mathcal{N}_k,x)d\omega,$$

$$\Upsilon_{j,kl}(t|s,y,\mathcal{N}_k,x) := \int_{\pi(s,y)}^{t} \lambda_j(\omega)\delta_{j,q}(\omega|\min\{s,y\},\mathcal{N}_q,x)\delta_{j,kl}(\omega|s,y,\mathcal{N}_{kl},x)d\omega,$$

$$\text{and } \Delta_{j,kl}(t|s,y,\mathcal{N}_{kl},x) := \int_0^t \delta_{j,kl}(\omega|s,y,\mathcal{N}_{kl},x)d\omega$$

exist and are finite, and iii) $\Delta_{j,k}(t|s,\mathcal{N}_k,x)$ *and* $\Delta_{j,kl}(t|s,y,\mathcal{N}_{kl},x)$ *are either cadlag or caglad in* s *and in* (s,y)*, respectively.*

Assumption A.1 states that there has to be sufficient variation of the covariate effects for each member. A sufficient condition for this assumption to be true is the existence of a continuous group-level characteristic and continuity of the function ϕ_j. It also imposes some innocuous normalization. Assumption A.2 is not restrictive as it allows for several parametric choices for the baseline hazard. Additionally, it normalizes the integrated baseline hazard for some particular value. Assumption A.3 is common in the analysis of the mixed proportional hazard model (Elbers and Ridder, 1982) and is needed to ensure identification[14]. Finally, Assumption A.4 give some (rather) weak finiteness conditions about the underlying interaction effects functions.

Let the statement $d_j = \infty$ imply that the corresponding subject never enters the risk process.[15] Define also $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$.

Proposition 1 *Let* $d_1 = 0, (d_2, d_3) \in \{\bar{\mathbb{R}}_+^2 : d_3 \geq d_2\}$*. Under Assumptions A.1-A.4, the set of functions* $\{\Lambda_j, \phi_j, \Delta_j, \Delta_{j,kl} : j,k,l = 1,2,3,\ k \neq j \neq l, k < l\}$ *and* G *in* MODEL A *are identified from the joint distribution of* $\{T_1, T_2, T_3\}|\{d, x\}$*.*

So far we have considered the case of varying entry dates across members and groups. In our empirical application this reflects the fact that siblings usually pass a fixed threshold

[14]Ridder and Woutersen (2003) discuss identification of the conventional mixed proportional hazard model by not imposing any conditions on the first moment of the unobserved term. We do not consider this case as it would be beyond the scope of this paper.

[15]Note that, for a maximal group size of M, all groups in the sample of size $J < M$ may be expressed by setting $d_{J+1} = ... = d_M = \infty$.

age, after which they become at risk of using drugs, at different calendar dates. It should be pointed out here, that we do not exploit full variation in d_j across members here.[16] Instead we only exploit variation across entry dates following a certain order $0 \leq d_2 \leq d_3$. The first born sibling is never born after the second and so forth.

On the one hand, different entry dates complicate the identification of Model A. With different entries the time until the first transition occurs within a group, can not be expressed in terms of a competing risk model for which standard identification results exist (see Heckman and Honoré, 1989; Abbring and Van den Berg, 2003a). On the other hand, varying entry dates provide an additional source of exogenous variation which allows to reduce the required variation in covariate effects to one dimension (see Assumption A.1). In the following subsection we discuss the special case of a common entry date for all members within a group.

2.2 Common entry dates

With some parallel-spell data, all group members enter the risk process at the same calendar date $d_1 = d_2 = d_3 = 0$. If, for example, a new product is introduced to a market, each member of a peer group becomes at risk of purchasing the new product at the same point in time. Similarly, a market specific shock, can be seen as a starting point after which each firm in the market is at risk to default. We first replace Assumption A.1 with Assumption A*.1.

Assumption A*.1 *The function $\phi_j : \mathbb{X} \to (0, \infty)$ is continuous with $\phi_j(x^*) = 1$ for some $x^* \in \mathbb{X}$, and $j = 1, 2, 3$. Moreover, the vector-valued mapping $(\phi_1(x), \phi_2(x), \phi_3(x); x \in \mathbb{X})$ contains a nonempty open subset of $\mathbb{R}^3_+$.*

Assumption A*.1 requires sufficient variation of the covariate effects across the three competing exit durations. It is analogous to one of the required assumptions in Abbring and Van den Berg (2003). Obviously, Assumption A*.1 is a bit stronger than Assumption A.1. Making use of a stronger requirement stems from the fact that in the case of common entry dates we cannot exploit variation in the timing of entry at the risk process for the group members.

[16] The case of full variation in entry dates across members, that is, when $d_1, d_2, d_3 \in \mathbb{R}_+ \cup \{\infty\}$, is a straightforward extension of Model A.The identification of the corresponding model is trivial by making use of Proposition 1.

Proposition 2 *Let $d_1 = 0, d_2 = 0, d_3 = 0$. Under Assumptions A^*.1,A.2,A.3, A.4, the set of functions $\{\Lambda, \Delta_{j,k}, \Delta_{j,kl} :,\ k \neq j \neq l, k < l\}$ and G in* Model A *are identified from the joint distribution $\{T_1, T_2, T_3\}|\{x\}$.*

A simple two-spell version of Model A with $d_1 = 0, d_2 = 0$ is formally introduced in Abbring and Heckman (2007) where an identification strategy is suggested by the authors.

2.3 Relaxing the proportionality assumption

In this subsection, we consider a set of conditions under which the proportionality assumption in Model A may be dropped. For this purpose we require some of the covariates to vary over time. More precisely, consider the covariate process $\chi_j : \mathbb{R}_+ \rightarrow \mathbb{X} \subseteq \mathbb{R}^\mu$ which is defined as follows $\chi_j(t) := (x_j^{'}(t)\ x_{-j}^{'}(t)\ x_g^{'}(t))^{'}$, where $x_{-j}(t)$ refers to the row vector of the individual characteristics of all members except for the $j-$th member and $x_g(t)$ holds all group-specific characteristics. Following Brinch (2008), we denote by $\mathcal{P}_\chi \subset \mathbb{R}_+ \times \mathbb{X}$ a family of this type of paths. We study the following multiple-spell duration model

Model B *Transition hazard of duration T_j given $T_k, T_l,\ d,\ \chi_j(t)$ and V_j*

$$\theta_j(t|T_k, T_l, d, \chi_j(t), V_j) = \tilde{\lambda}(t, \chi_j(t))\delta_j(t|T_k, T_l, d, \chi_j(t))V_j$$

with social interaction effect functions

$$\delta_j(t|T_k, T_l, d, \tilde{\chi}_j(t)) = \delta_{j,k}(t|T_{jk}, \mathcal{N}_k, \chi_j(t))^{I_{j,k}(t)}\delta_{j,l}(t|T_{jl}, \mathcal{N}_l, \chi_j(t))^{I_{j,l}(t)}\delta_{j,kl}(t|T_{jk}, T_{jl}, \mathcal{N}_{kl}, \chi_j(t))^{I_{j,kl}(t)},$$

where $\mathcal{N}_k, I_{j,k}(t), \mathcal{N}_{kl}$, and *$I_{j,kl}(t)$ have the same interpretation as in* Model A with *$j, k, l = 1, 2, 3$ such that $k \neq j \neq l$ and $k < l$.*

Assumption B.1 *The function $\tilde{\lambda} : \mathbb{R}_+ \times \mathbb{X} \rightarrow (0, \infty)$ is measurable and the integrated generalized baseline hazard rate $\tilde{\Lambda}(t, \chi_j) := \int_0^t \tilde{\lambda}(\omega, \chi_j(\omega))d\omega$ exists and is finite for all $t > 0$ and $\chi_j \in \mathcal{P}_\chi$, and $j = 1, 2, 3$.*

Assumption B.2 *There are two distinct covariate paths $\chi_1 \in \mathcal{P}_\chi$ and $\xi_1 \in \mathcal{P}_\chi$ such that $\chi_1(t) = \xi_1(t)$ for some $t \in (t_a, t_b)$ with $t_a < t_b$ and $\tilde{\Lambda}(t_a, \chi_1) \neq \tilde{\Lambda}(t_a, \xi_1)$.*

Assumption B.3 *The G is such that does not depend on x and d .*

Assumption B.4 *For $j, k, l = 1, 2, 3$ such that $k \neq j \neq l$ and $k < l$. Let $q = \arg\min_{k,l}\{T_{jk}, T_{jl}\}$ and $\pi(s, y) = \max\{0, \min\{s, y\}\}$. The functions $\delta_{j,k} : \mathbb{R}_+ \times \mathbb{R} \times \{1, 2, 3\} \times \mathbb{X} \to (0, \infty)$, and $\delta_{j,kl} : \mathbb{R}_+ \times \mathbb{R}^2 \times \{1, 2, 3\} \times \mathbb{X} \to (0, \infty)$ are measurable, ii) the quantities*

$$\Upsilon_{j,k}(t|s, \mathcal{N}_k, \chi_j) := \int_{\max\{0,s\}}^{t} \tilde{\lambda}(\omega, \chi_j(\omega))\delta_{j,k}(\omega|s, \mathcal{N}_k, \chi_j(\omega))d\omega,$$

$$\Delta_{j,k}(t|s, \mathcal{N}_k, \chi_j) := \int_0^t \delta_{j,k}(\omega|s, \mathcal{N}_k, \chi_j(\omega))d\omega,$$

$$\Upsilon_{j,kl}(t|s, y, \mathcal{N}_{kl}, \chi_j) := \int_{\pi(s,y)}^{t} \tilde{\lambda}(\omega, \chi_j(\omega))\delta_{j,q}(\omega| \min\{s, y\}, \mathcal{N}_q, \chi_j(\omega))\delta_{j,kl}(\omega|s, y, \mathcal{N}_{kl}, \chi_j(\omega))d\omega,$$

$$\text{and } \Delta_{j,kl}(t|s, y, \mathcal{N}_{kl}, \chi_j) := \int_0^t \delta_{j,kl}(\omega|s, y, \mathcal{N}_{kl}, \chi_j(\omega))d\omega$$

exist and are finite, and iii) $\Delta_{j,k}(t|s, \mathcal{N}_k, \chi_j)$ and $\Delta_{j,kl}(t|s, y, \mathcal{N}_{kl}, \chi_j)$ are either cadlag or caglad in s and in (s, y), respectively.

Assumption B.1 deals with measurability and finiteness conditions of the (integrated) generalized baseline hazard. Assumption B.2 ensures that there exist two different covariate paths which agree on an open interval. Note that the latter can be satisfied by just considering a single covariate which will meet the condition of Assumption B.2. In contrast to Assumption A.3, Assumption B.3 does not impose any conditions on the first moment of the unobserved terms. This is due to the presence of time-varying covariates (Heckman and Taber 1994; Brinch, 2008). Finally, Assumption B.4 is similar to Assumption A.4 and is concerned with finiteness conditions of the underlying functions.

Proposition 3 *Let $d_1 = 0, (d_2, d_3) \in \{\bar{\mathbb{R}}_+^2 : d_3 \geq d_2\}$. Under Assumptions B.1-B.4, the set of functions $\{\tilde{\Lambda}, \Delta_{j,k}, \Delta_{j,kl} : j, k, l = 1, 2, 3,\ k \neq j \neq l, k < l\}$ and G in* Model B *are identified from the joint distribution of $\{T_1, T_2, T_3\}|\{d, x\}$.*

Note that, in contrast to Model A, the finiteness of the first moment of the unobserved terms is not necessary due to the presence of time-varying covariates (Heckman and Taber 1994; Brinch, 2008).

As in the case with different entry dates, we can relax the proportionality assumption in the setting with common entry dates as well. In particular, we introduce the covariate process

$\zeta_j : \mathbb{R}_+ \to \mathbb{Z} \subseteq \mathbb{R}^{\bar{\mu}}$ which is obtained as follows $\zeta_j(t) := (x_j^{'}(t) \quad x_g^{'}(t))'$, and the family of such processes $\mathcal{P}_\zeta \subset \mathbb{R}_+ \times \mathbb{Z}$. Note that $\bar{\mu} < \mu$ as the process $\zeta_j(t)$, in contrast to the process $\chi_j(t)$, does not include the characteristics of members other than j. Consider the following assumptions.

Assumption B*.5 *It holds* $\tilde{\lambda}(t, \chi_j) = \tilde{\lambda}(t, \zeta_j)$ *for all* $t > 0, \chi_j \in \mathcal{P}_\chi,\ \zeta_j \in \mathcal{P}_\zeta, j = 1, 2, 3.$ [17]

Assumption B*.6 *The vector-valued mapping* $(\tilde{\Lambda}(t, \zeta_1), \tilde{\Lambda}(t, \zeta_2), \tilde{\Lambda}(t, \zeta_3); \zeta_1, \zeta_2, \zeta_3 \in \mathcal{P}_\zeta, t \in \mathbb{R}_+)$ *contains a nonempty open subset of* $\mathbb{R}_+^3$.

Assumption B*.5. implies that the generalized baseline hazard for each member does not depend on the individual characteristics of the other group members. Moreover, Assumption B*.6 imposes the condition that the three integrated generalized baseline hazard can independently of each other vary on $\mathbb{R}_+^3$. A sufficient condition for this statement to be true is the existence of a certain member-specific characteristic which affects only the member directly but not the other group members.

Proposition 4 *Let* $d_1 = 0, d_2 = 0, d_3 = 0$. *Under Assumptions B.1,- B.4, B*.4,- B*.6, the set of functions* $\{\tilde{\Lambda}, \Delta_{j,k}, \Delta_{j,kl} : j, k, l = 1, 2, 3,\ k \neq j \neq l, k < l\}$ *and* G *in* Model B *are identified from the joint distribution of* $\{T_1, T_2, T_3\}|\{d, x\}$.

3 Empirical Application

In the following we present our empirical Application. First, we introduce our data set, then we discuss the estimation method and finally present our results.

3.1 Data

In our empirical study we use data from the National Longitudinal Survey of Youth 1979 (see The NLSY79, 2005 for an introduction). The NLSY79 survey was established in an effort

[17] To keep the notation simple we have chosen to abuse a bit the notation. Specifically, we keep on using the same notation for the extended baseline hazard, $\tilde{\lambda}$, even if this does not depend on the individual characteristics of the other members as it happens with the case of non-common entruy dates.

to generate a representative sample of young men and women aged 14 to 21 living in the United States. Respondents are drawn from cohorts 1957 to 1964 and for each respondent, all individuals aged 14 to 21 living in the same household at the time of the first round in 1979, were also included in the survey.[18] 12,686 respondents were included this way living in 7,490 unique households. We restrict attention to single-respondent households and households with more than one respondent where the respondents are siblings (blood-related and not blood-related) and grew up together in their parents home.[19] We observe 8,684 respondents in 5,810 unique households satisfying these criteria, of which 1,549 comprise two, 516 three and 153 four to six siblings.

In the 1984 survey three separate questions were asked addressing first-time marijuana use. The respondents were asked in which year and which month they started using marijuana/hashish for the first time in their life. 5578 respondents report month and year and 3,723 have never used up to the interview date in 1984. Based on this and using information on birth dates of the respondents, we can construct the durations until first time drug use after passing the threshold age of 7 for each household member. For the respondents who have never used, the durations are censored at the time of the interview date. For 178 respondents no transition times are reported (174 respondents answer the question with "Don't know" and 4 were not interviewed or refused to respond) In addition to the question on first-time drug use, a monthly time-line of marijuana use for the past 4,5 years was established in July 1984. Furthermore, in surveys 1988, 1990, 1992, and 1994-2008 respondents are asked how old they were when they first used marijuana. Combining the information of these three questions provides a detailed retrospective picture on drug-use behavior. From this we construct an index measuring the degree of uncertainty in the responses due to inconsistencies in the answers. This index may be used in a sensitivity analysis.

We combine the detailed information on monthly marijuana-use from Jan 1979 to July 1984 with annual information on the amount and frequency for all relevant survey years.

[18]This way the households are not complete, in the sense that only the siblings from cohorts 1957 to 1964 are included as respondents in the survey. We will refer to these incomplete groups as households from now on.

[19]In the majority of all households selected this way, the siblings grew up living with both biological parents. We can observe the time when individuals leave their parents home and the reason for this move (e.g. divorce of the parents). In the analysis of social interaction effects we account for this by not allowing interaction effects at calendar dates where the members do not live in the same household.

Based on this, we can select the cases where a first-time transition is followed by a long-term change in drug-use behavior.

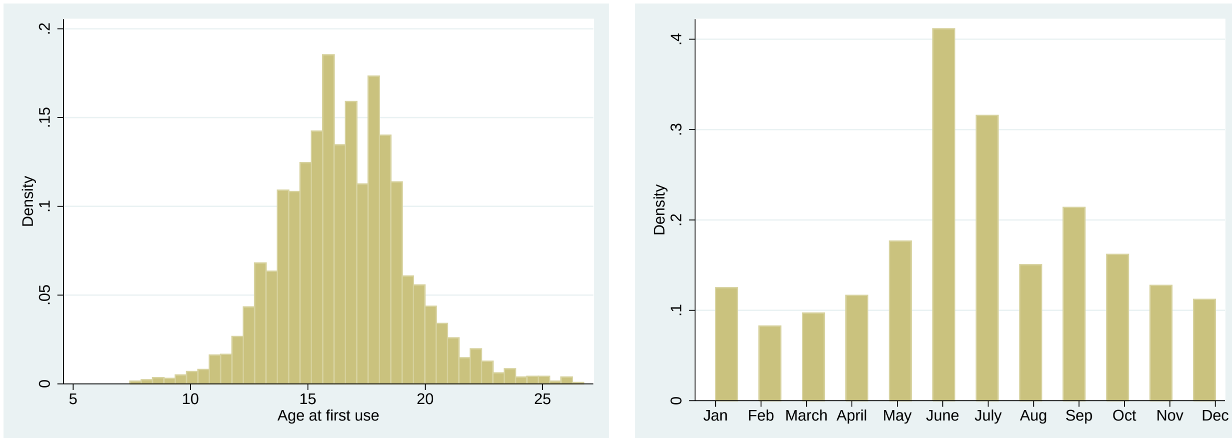

Figure 2: Pooled over all household members: (left) Distribution of age at transition; (right) Distribution of the month of transition.

The resulting distribution of transition times pooled over all household members is presented in Figure 2. The left figure shows the distribution of age at transition. Before the age of seven, only very few transitions occur. We drop those cases from the sample[20] and choose the age of seven as the threshold age after which siblings become at risk of using drugs. The majority of transitions occurs between the age of 14 and 18. The right figure depicts the distribution of the month at transition. There is a strong peak over the summer months June and July in which American teenagers often go to summer camp and/or spend a lot of time outside. In September the number of transitions increase again. At this time teenagers enter a new year in high school and are exposed to many new influences. In our empirical analysis we control for the different effects by adding time-varying dummies for each month to the vector of covariates.

Figure 3 shows the estimated transition (baseline) hazards from a single spell Cox proportional hazard model. There is a substantial difference between the first marijuana use times of oldest and youngest siblings in the households. Younger siblings transition at an earlier age compared to their older siblings. This effect could be driven by observable or unobservable

[20]These cases are most likely a result of the measurement error caused by the retrospective nature of the fist-time drug use question.

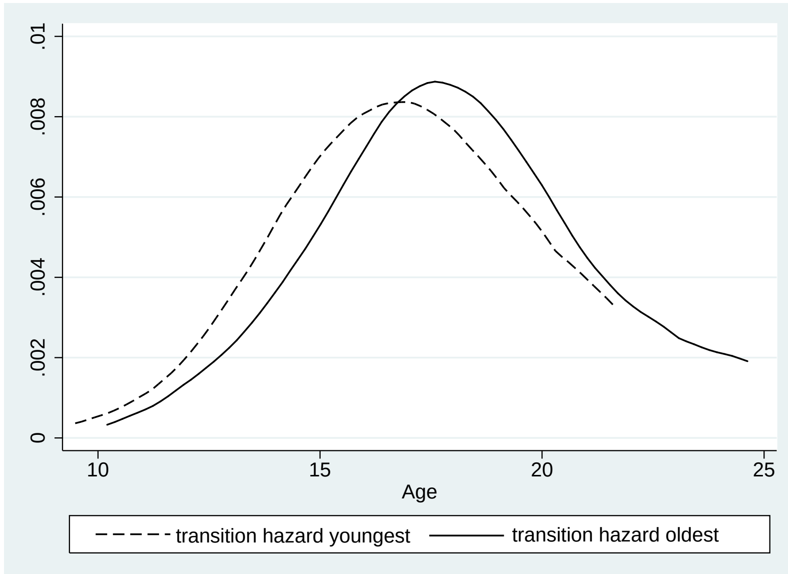

Figure 3: Estimated baseline hazards of a single spell Cox proportional hazards model using data from the NLSY79 cohorts 1957-64 of households with at least three siblings on first-time marijuana use of the youngest (dashed) and oldest (solid) sibling in the household.

characteristics such as the cohort or character traits which are different for the oldest compared to the youngest siblings. An alternative explanation is the existence of positive social interaction effects. Younger siblings experience the transitions of their older siblings making them more likely to transition at an early age.

3.2 Maximum Likelihood Estimation

Model A provides a general framework of a multiple-spell model with interaction effects which allows to specify a variety of models fitting different applications. In order to estimate a model using data on first-time marijuana use, we specify the functional forms of λ_j, ϕ_j, δ_j and G. This way the semi-parametric form of Model A is reduced to a model with a finite set of parameters. The resulting parametric model can be estimated using standard maximum likelihood methods.

Figure 4 (dashed line) shows the estimated baseline hazard of a Cox proportional hazard model with a shared frailty term on the household level and a basic set of covariates. We use the log-logistic density function to approximate this shape in the estimation of our full model.

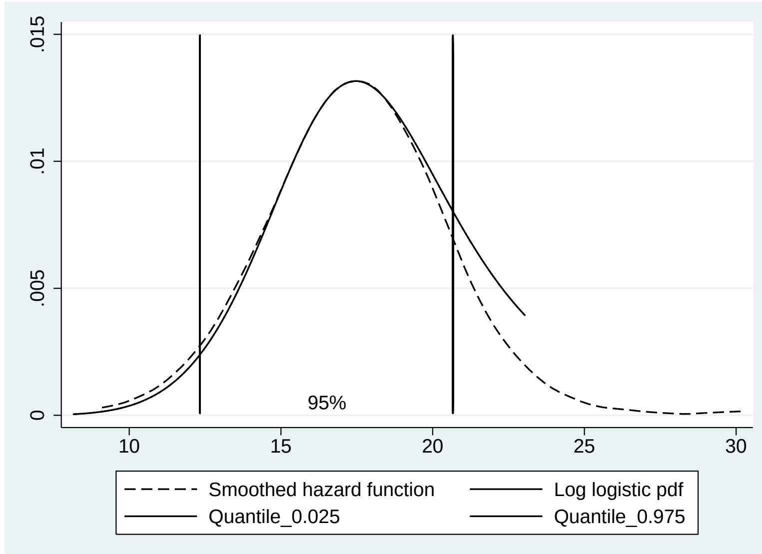

Figure 4: Estimated baseline hazard functions using data from the NLSY79 cohorts 1957-64 on first-time marijuana use: (dashed) Cox proportional hazard model; (solid) Parametric proportional hazard model using a Log-logistic probability density function for the baseline hazard. Both models are estimated with a basic set of covariates a shared frailty term on the household level.

This function has a positive range and is able to approximate the shape of the baseline hazard estimated by the more flexible Cox model (see Figure 4). In the main model specification we assume proportionality of the covariate effects (Model A). This leads to the following baseline and regression component function and the corresponding integral of this function for sibling j in household i at duration t (counted in months)

$$\lambda_j(t)\ \phi_j(x_{ij}(t)) = \frac{\alpha_{2,j}}{\alpha_{1,j}}(\frac{t}{\alpha_{1,j}})^{\alpha_{2,j}-1}(1+(\frac{t}{\alpha_{1,j}})^{\alpha_{2,j}})^{-2}\ e^{\beta_{0,j}+\beta' x_{ij}(t)}$$

$$\tilde{\Lambda}_j(t, x_{ij}(t)) = \sum_{\tau=1}^{t}[(1+(\frac{\tau}{\alpha_{1,j}})^{-\alpha_{2,j}})^{-1} - (1+(\frac{\tau-1}{\alpha_{1,j}})^{-\alpha_{2,j}})^{-1}]\ e^{\beta_{0,j}+\beta' x_{ij}(\tau-1)}$$

with $\alpha_{q,j} = \alpha_{q,oldest}$ for $j = 1$ and $\alpha_{q,j} = \alpha_{q,young}$ for $j > 1$, $q = 1, 2$.

Further we specify the interaction effect δ_j with several multiplicative terms, each representing the influence of an experienced transition of a sibling. Further an additional multiplicative term accounts for the number of transitions experienced within the household up to this point. For a sibling j living in a household i with J_i members at time t this yields

$$\delta_j(t|T_{i,-j}, x_{ij}(t)) = \prod_{k\in -j} \delta_{j,k}(t|T_{i,jk}, x_{ij}(t))^{I(t>T_{i,jk})}\quad e^{\gamma_{count}\sum_{l\in -j} I(t>T_{i,jl})}$$

$$\text{with} \qquad \delta_{j,k}(t|T_{i,jk}, x_{ij}(t)) = exp(\ \gamma_k + \gamma_x^{'} x_{ij}(t) + \gamma_{x_{int}}^{'}(x_{ij}(t) \times x_{ik}(t)\ )$$

with $T_{i,-j} := \{T_{i,jk} : k \in -j\}$ and $-j := \{k \in J_i : k \neq j\}$ and $\gamma_k = \gamma_{oldest}$ for $j = 1$ and $\gamma_k = \gamma_{young}$ for $k > 1$.

We capture unobserved heterogeneity in the transition hazards by two additive components. The term V_j of sibling j of household i is given by

$$V_{ij} = V_i^{sh} + V_{ij}^{ind}$$

Here, the random terms V_i^{sh} and V_{ij}^{ind} are independently drawn from distributions G^{sh} and G^{ind} with the mean of V_{ij} normalized to 1. The former term captures unobserved heterogeneity of the hazard rates across households whereas the latter reflects unobserved heterogeneity within households across different members. We assume that V_i^{sh} can attain two values m_1^{sh} and m_2^{sh} with $P(V_i^{sh} = m_1^{sh}) = p^{sh}$, representing households with high or low susceptibility to drug use. Similarly, V_{ij}^{ind} can attain two values m_1^{ind} and m_2^{ind} with $P(V_i^{ind} = m_1^{ind}) = p^{ind}$. This way the distribution of V_{ij}, which is the sum of V_i^{sh} and V_{ij}^{ind}, has four mass-points. Note that the term V_i^{sh} which is shared across members of the same household generates a correlation between terms V_{ij} and V_{ik} $\rho_{i,jk} = \frac{\sigma_{sh}^2}{\sigma_{sh}^2 + \sigma_{ind}^2}$, where $\sigma_{sh}^2 = Var(V^{sh})$ and $\sigma_{ind}^2 = Var(V^{ind})$.

From this we can construct the hazard rate and survival function of each household member $j \in J_i$ given the transition times of the other members $k \in -j$

$$\theta_j(t|\{T_{i,-j}\}, x_{ij}(t), V_{ij}) = \frac{\frac{\alpha_{2,j}}{\alpha_{1,j}}(\frac{t}{\alpha_{1,j}})^{\alpha_{2,j}-1}}{(1 + (\frac{t}{\alpha_{1,j}})^{\alpha_{2,j}})^2}\ e^{\beta_{0,j} + \beta^{'} x_{ij}(t)} \prod_{k \in -j} \delta_{j,k}(t|T_{i,jk}, x_{ij}(t))^{I(t > T_{i,jk})}\ \ e^{\gamma_{count} \sum_{l \in -j} I(t > T_{i,jl})}\ \ V_{ij} \tag{4}$$

$$S_j(t|\{T_{i,-j}\}, x_{ij}(t), V_{ij}) = exp(-\sum_{l \in -j} I(T_{i,jk} > 0)[\tilde{\Lambda}_j(T_{i,jl}, x_{ij}(t)) - \tilde{\Lambda}_j(\max_{k \in -j_l}\{0, T_{i,jk}\}, x_{ij}(t))]\ \delta_j(T_{i,jl}|\{T_{i,-j}\}, x_{ij}(t))\ V_{ij}\ ) \tag{5}$$

with $\{-j_l\} := \{k \in J_i : k \neq j \wedge T_{i,jk} < T_{i,jl}\}$.

In the following we denote the transition durations of each household i by the vector of random variables $T_i = (T_{i1}\ ...\ T_{iJ_i})$ and their realizations by $t_i = (t_{i1}\ ...\ t_{iJ_i})$. The durations within each household are observed only up to a common calendar time at which the interview is conducted in 1984. We denote the resulting vector of censoring points as $c_i = (c_{i1} \ldots c_{iJ_i})$.[21] With this information we can construct the likelihood contribution of a household i with J_i members.

$$
\begin{aligned}
&\text{Ł}(t_i, c_i, x_i; \alpha, \beta, \gamma, m, p) \\
&= \int_0^\infty (\prod_{j \in J} \int_0^\infty \theta_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), V_{ij})^{I(c_j=0)} S_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), V_{ij}) dG^{ind} \;) dG^{sh} \\
&= \sum_{q=1}^{2} \sum_{q_1=1}^{2} \ldots \sum_{q_{J_i}=1}^{2} \prod_{j \in J_i} \theta_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), m_q^{sh} + m_{q_j}^{ind})^{I(c_j=0)} S_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), m_q^{sh} + m_{q_j}^{ind}).
\end{aligned}
\tag{6}
$$

3.3 Results

We estimate our model of first time use of marijuana based on the likelihood function presented in (4) - (6) at the end of Section 3.2. In our analysis we use data on 669 households with at least three siblings growing up together. The results of three different model specifications are reported in Table 1. Model I represents a simple model with covariates and a basic specification of social interaction effects but without accounting for unobserved characteristics (no Correlated Effects: $\sigma^2_{sh} = \sigma^2_{ind} = 0$). Two parameter estimates for the social interaction effect functions γ_{oldest} and $\gamma_{younger}$ are reported ($\gamma_x = \gamma_{x_{int}} = 0$). The parameter γ_{oldest} represents how the transition hazard of a sibling is affected if he/she experiences that the oldest sibling in the household starts using marijuana. $\gamma_{younger}$ measures the effect if one of the younger siblings starts using drugs. In this simple model, we find highly significant and strongly positive estimates of these parameters. However, Model II reveals that the estimates in Model I pick up a dependence between group members generated by unobserved characteristics (Correlated effects). When we account for Correlated Effects in Model II, we still find a highly significant positive effect of a transition of the oldest sibling in the household but do not find a significant effect for the transition of a younger sibling. In Model III we allow for additionally flexibility of the social interaction effect functions. We find that females

[21] Note that, household members are censored at the same calendar time. The resulting censoring durations $c_{i1}, ..., c_{iJ_i}$ may differ due to different entry dates of the members (age difference between siblings).

are more strongly influences by a transition of their fellow siblings than males. Furthermore, we do not find evidence for an effect of family net income on the strength of social interactions within households. The last two parameters reported for Social Interaction Effects reflect the estimated effects of a dummy which has value one if the sibling who starts using drugs and the sibling how is affected by this transition are both female (or both male). We do not find a significant effect for the same gender.

The estimated probabilities and mass points described in Section 3.2 imply variances σ^2_{sh}, σ^2_{ind} of the two distributions G^{sh}, G^{ind} and correlation ρ_{jk} between the unobserved heterogeneity terms of two group members V_{ij} and V_{ik}. The parameters are reported in the section Correlated Effects in Table 1. We find evidence for unobserved heterogeneity across groups ($\sigma^2_{sh} \approx 0.1$) but not within groups ($\sigma^2_{ind} \approx 0.01$) in Models II and III. This implies a high correlation of V_{ij} and V_{ik} between two group members.

In this empirical section we find evidence for the fact that the oldest sibling in a household influences his younger siblings in terms of his marijuana use. We however do not find evidence for an effect of a transition of a younger sibling. Females are more strongly influenced by the drug use behavior of their siblings than males. Furthermore, besides observable characteristics and social interaction effects, unobserved characteristics shared among siblings explain part of the dependence in the drug use behavior.

	Model I		Model II		Model III	
Variable	estimate	st.error	estimate	st.error	estimate	st.error
Covariates:						
Oldest sibling	.281***	(.089)	.237***	(.098)	.266***	(.099)
Female	-.543***	(.064)	-.619***	(.071)	-.736***	(.09)
Birth year	-.015	(.019)	.008	(.023)	.037	(.031)
Number Siblings	-.384***	(.122)	-.457***	(.147)	-.225	(.194)
Family net income	-.019	(.036)	-.028	(.044)	-.009	(.045)
Father employed	-.129**	(.064)	-.122	(.081)	-.211*	(.109)
Poverty	-.04	(.076)	-.036	(.095)	-.037	(.093)
Both parents in HH	.433***	(.157)	.475***	(.186)	.516***	(.183)
School attendance	.004	(.007)	-.002	(.008)	-.001	(.008)
White	.289***	(.063)	.386***	(.082)	.376***	(.081)
Urban	.225***	(.073)	.238***	(.091)	.252***	(.089)
Social Interaction Effects:						
Sibling transitioning:						
Oldest sibling γ_{oldest}	.634***	(.08)	.347***	(.113)	.252*	(.146)
Younger sibling γ_{young}	.386***	(.056)	.132	(.085)	-.019	(.128)
Sibling affected:						
Female	-	-	-	-	.208**	(.104)
Birth year	-	-	-	-	-.033*	(.02)
Number Siblings	-	-	-	-	-.254	(.156)
Family net income	-	-	-	-	.005	(.028)
Same characteristics:						
Female	-	-	-	-	.119	(.091)
Male	-	-	-	-	.021	(.095)
Correlated Effects:						
Shared term σ^2_{sh}	-	-	.118	.	.105	.
Indiv. term σ^2_{ind}	-	-	.010	.	.008	.
Correlation ρ_{jk}	-	-	.914	.	.921	.
Month dummies	YES		YES		YES	
Households $\geq$ 3 sib	669		669		669	
Time periods	325		325		325	
LogLikelihood	-10179.3		-7381.5		-7372.4	

Table 1: Estimation of the specification of Model A described in Section 3.2 using data from the NLSY79 on first time drug use of siblings in 669 American households in which at least three siblings grew up together. Estimates with *, ** or *** reflect a 0.1, 0.05 or 0.01 significance level.

4 Conclusion

Our empirical results suggest that the oldest sibling has a distinct social role in the household i.e. his behavior has a strong influence on the younger siblings, but not vice versa. This reveals there can be strong asymmetries across different group members in terms of their potential influence on others. Our approach can be used to identify such key members within a group, and can predict the development of social multiplier effects over time. This allows predicting the impact of public policies, depending on which members are initially targeted.

Our approach provides an alternative to interdependent utility equilibrium models in studying social interactions from transition data. We argue that in applications such as substance use of teenagers, a transition of a peer can have the characteristic of an unanticipated shock and may directly alter the behavior of other group members. Our approach exploits the information on the exact timing of actions within a group, whereas standard approaches do not make use of this information. This may be driven by the limitation of yearly survey data, which is primarily used in studies of social interactions. However, administrative data and data on interactions online constitute an increasingly important data source, providing very detailed information on the timing of actions. Being able to exploit this information may become increasingly valuable.

Appendix

Notation

Before proceeding, we will introduce some notation and conventions which will be used throughout the Appendices. The symbol G with some (double) subscript will refer to the corresponding marginal or bivariate distribution. For instance, G_{12} denotes the bivariate distribution of $(V_1\ V_2)'$. No superscript at G denotes, as already adopted in the main text, the full trivariate distribution of $(V_1\ V_2\ V_3)'$. Also, we will use the generic symbol $\mathcal{L}$ to denote the Laplace Transform of some probability measure. The (double) superscript at $\mathcal{L}$ will indicate the corresponding (mixed) partial derivative. To give an example, $\mathcal{L}_G^{(23)}$ denotes the mixed partial derivative with respect to the second and third argument of the Laplace Transform of G. Finally, let $\bar{\mathbb{D}} := \{d_1 = 0, (d_2, d_3) \in \bar{\mathbb{R}}_+^2 : d_3 \geq d_2\}$, $\mathbb{D} := \{d_1 = 0, (d_2, d_3) \in \mathbb{R}_+^2 : d_3 \geq d_2\}$, $\mathbb{D}_\infty := \{d_1 = 0, d_2 \in \mathbb{R}_+, d_3 = \infty\}$, and $\mathbb{D}_{2\infty} := \{d_1 = 0, d_2 = \infty, d_3 = \infty\}$.

For the proof of the propositions we will utilize certain subsurvival functions. More precisely, for $t > 0, x \in \mathbb{X}$, $d \in \bar{\mathbb{D}}$, and $j = 1, 2, 3$,

$$Q_{T_j}(t|d, x) := \mathbb{P}(T_j > t, T_j + d_j < \min_{k \in \{1,2,3\} \neq j} (T_k + d_k)|d, x).$$

In addition, for $t_1, t > 0$, $x \in \mathbb{X}$, and $j = 2, 3$,

$$Q_{T_1}(t_1, t|d, x) := \begin{cases} \mathbb{P}(T_1 > t_1, T_2 > t, T_1 < T_2 + d_2|d, x) & \text{if } d \in \mathbb{D}_\infty, \\ \mathbb{P}(T_1 > t_1, T_2 > t + d_3 - d_2, T_3 > t, T_1 < \min_{k \in \{2,3\}}(T_k + d_k)|d, x) & \text{if } d \in \mathbb{D}. \end{cases}$$

$$Q_{T_1,T_j}(t_1, t|d, x) := \mathbb{P}(T_1 > t_1, T_2 > t + d_3 - d_2, T_3 > t, T_1 < T_j + d_j < T_k + d_k|d, x) \text{ if } d \in \mathbb{D}.$$

Finally, for $t_1, t_j, t_k > 0, x \in \mathbb{X}$, and $j, k = 2, 3$ such that $j \neq k$,

$$Q_{T_1,T_j,T_k}(t_1, t_j, t_k|d, x) := \mathbb{P}(T_1 > t_1, T_j > t_j, T_k > t_k, T_1 < T_j + d_j < T_k + d_k|d, x) \text{ if } d \in \mathbb{D}.$$

Proof of Proposition 1

The proof of Proposition 1 consists of three main steps. The first step describes the identification of the integrated baseline hazards, the regressor functions, and the distribution function of the unobserved heterogeneity terms. The second step deals with the identification of the interaction effects caused by the first exit. Finally, the third step is concerned with the identification of the interaction effects caused by the second exit.

Identification of the set of functions $\{\Lambda_j, \phi_j : j = 1, 2, 3\}$ and G. For all $t > 0$, $x \in \mathbb{X}$, and $d \in \mathbb{D}_{2\infty}$, we have

$$\mathbb{P}\left[T_1 > t|d, x\right] = \mathcal{L}_{G_1}(\phi_1(x)\Lambda_1(t)). \tag{A.1}$$

Following analogous steps to Elbers and Ridder (1982), we achieve identification of ϕ_1, G_1, and Λ_1.

Next, we identify ϕ_2 and Λ_2. For almost any $t > 0$, $x \in \mathbb{X}$, and $d \in \mathbb{D}_\infty$, we obtain

$$\frac{\partial}{\partial t} Q_{T_2}(t|d, x) = \phi_2(x)\lambda_2(t)\mathcal{L}^{(2)}_{G_{12}}(\phi_1(x)\Lambda_1(t + d_2), \phi_2(x)\Lambda_2(t)), \tag{A.2}$$

It is straightforward, by Assumption B.3, to check that

$$\lim_{t\to 0, d_2\to 0} \left[\frac{\partial}{\partial t} Q_{T_2}(t|d, x) \Big/ \frac{\partial}{\partial t} Q_{T_2}(t|d, x^*)\right] = \phi_2(x), \tag{A.3}$$

which leads to identification of ϕ_2. For any $t > 0$, $x \in \mathbb{X}$, and $d \in \mathbb{D}_\infty$,

$$\mathbb{P}\left[\bigcap_{j=1}^{2} (T_j + d_j > t + d_2)\Big| d, x\right] = \mathcal{L}_{G_{12}}(\phi_1(x)\Lambda_1(t + d_2), \phi_2(x)\Lambda_2(t)). \tag{A.4}$$

We let $t = t^*$ and thus we can trace out $\mathcal{L}_{G_{12}}$ on an open subset of $\mathbb{R}^2_+$ by varying appropriately d_2 and x. Given that $\mathcal{L}_{G_{12}}$ is real analytic function (Abbring and Van den Berg, 2003a), we identify $\mathcal{L}_{G_{12}}$ (and consequently G_{12}) on $\mathbb{R}^2_+$. Then, employing the relation (A.4), we identify Λ_2. The identification of ϕ_3, Λ_3, and G follows the same line of argument as that in identification of ϕ_2, Λ_2, and G_{12}, and is consequently omitted. ■

For the second and third step note that for $t > 0$

$$\Delta_{j,k}(t|.) = \int_0^t \frac{\partial \Upsilon_{j,k}(\omega|.)}{\partial \omega} \left[\lambda_j(\omega)\right]^{-1} d\omega,$$

and

$$\Delta_{j,kl}(t|.) = \int_0^t \frac{\partial \Upsilon_{j,kl}(\omega|.)}{\partial \omega} \left[\lambda_j(\omega)\right]^{-1} \delta_{j,kl}(\omega|.) d\omega.$$

Hence, to identify $\Delta_{j,k}$ and $\Delta_{j,kl}$ it is sufficient to identify $\Upsilon_{j,k}$ and $\Upsilon_{j,kl}$, respectively.

Identification of the set of functions $\{\Delta_{j,k} : j, k = 1, 2, 3, \ j \neq k\}$**.** We begin with the identification of $\Delta_{2,1}$ and $\Delta_{3,1}$. Three different cases are possible: i) $0 < T_1 \leq d_2$, ii) $d_2 < T_1 \leq d_3$, and iii) $T_1 > d_3$. The identification methodology can be summarized as follows. We first identify $\Upsilon_{2,1}$ for the cases i) and ii), next we identify $\Upsilon_{3,1}$ for the cases i) and ii), and finally we jointly identify $\Upsilon_{2,1}$ and $\Upsilon_{3,1}$ for the case iii).

For almost all t_1 such that $0 < t_1 \leq d_2$, each $t > 0$, $d \in \mathbb{D}_\infty$, and $x \in \mathbb{X}$,

$$\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G_{12}}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t|t_1 - d_2, 1, x)). \tag{A.5}$$

By the first step, all the quantities on the right hand side are known except for the term $\Upsilon_{2,1}$. By exploiting the facts that $\mathcal{L}^{(1)}_{G_{12}}$ is strictly increasing in its arguments and that $\Upsilon_{2,1}(t|t_1 - d_2, 1, x)$ is either cadlag or caglad in $t_1 - d_2$ (Assumption A.4), we can identify $\Upsilon_{2,1}$ for the case i). Similarly, for almost every t_1 such that $d_2 < t_1 \leq d_3$, all $t > t_1 - d_2$, $d \in \mathbb{D}_\infty$, and $x \in \mathbb{X}$,

$$\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G_{12}}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)(\Lambda_2(t_1 - d_2) + \Upsilon_{2,1}(t|t_1 - d_2, 2, x))). \tag{A.6}$$

Identical arguments to the previous case give identification of $\Upsilon_{2,1}$ for the case ii).

Next, we proceed with the identification of $\Upsilon_{3,1}$ for the first two cases. More precisely, for almost all $0 < t_1 \leq d_2$, all $t > 0$, $d \in \mathbb{D}$, and $x \in \mathbb{X}$ we obtain

$$\begin{aligned}\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G}(&\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t + d_3 - d_2|t_1 - d_2, 1, x), \\ &\phi_3(x)\Upsilon_{3,1}(t|t_1 - d_3, 1, x)).\end{aligned} \tag{A.7}$$

Next, we note that for almost every $d_2 < t_1 \leq d_3$, all $t > 0$, $d \in \bar{\mathbb{D}}$, and $x \in \mathbb{X}$,

$$\begin{aligned}\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G}(&\phi_1(x)\Lambda_1(t_1), \phi_2(x)\left(\Lambda_2(t_1 - d_2) + \Upsilon_{2,1}(t + d_3 - d_2|t_1 - d_2, 2, x)\right), \\ &\phi_3(x)\Upsilon_{3,1}(t|t_1 - d_3, 2, x)).\end{aligned} \tag{A.8}$$

Recall that $\Upsilon_{2,1}$ has been identified for the two above cases. Then, the $\Upsilon_{3,1}$ can be uniquely determined for the corresponding cases.

Finally, we turn our attention to the case *iii*). Note that for almost all $t > 0, d \in \mathbb{D}$, $x \in \mathbb{X}$,

$$\lambda_j(t+\eta_j) = \frac{\partial Q_{T_j}(t|d,x)}{\partial t}\Big[\mathcal{L}_G^{(j)}(\phi_1(x)\Lambda_1(t+d_3), \phi_2(x)\Lambda_2(t+d_3-d_2), \phi_3(x)\Lambda_3(t))\phi_j(x)\Big]^{-1}, \tag{A.9}$$

where $j = 2, 3$, $\eta_2 = d_3 - d_2$, and $\eta_3 = 0$. For almost all $t_1 > d_3$, almost each $t > t_1 - d_3$, $d \in \bar{\mathbb{D}}$, $x \in \mathbb{X}$,

$$\begin{aligned}\lambda_j(t+\eta_j)\delta_{j,1}(t+\eta_j|t_1-d_j,3,x) = \Big[\mathcal{L}_G^{(1j)}(&\phi_1(x)\Lambda_1(t_1),\\ &\phi_2(x)(\Lambda_2(t_1-d_2)+\Upsilon_{2,1}(t+d_3-d_2|t_1-d_2,3,x)),\\ &\phi_3(x)(\Lambda_3(t_1-d_3)+\Upsilon_{3,1}(t|t_1-d_3,3,x)))\\ &\times\phi_1(x)\lambda_1(t_1)\phi_2(x)\Big]^{-1}\frac{\partial^2 Q_{T_1,T_j}(t_1,t|d,x)}{\partial t_1\partial t}.\end{aligned} \tag{A.10}$$

The rest of this part is analogous to the proof of Proposition 1 of Drepper and Effraimidis (2012). We fix t_1, x, d_2, and d_3. Define $\mathcal{H}_j(t) := \Lambda_j(t+\eta_j)$ and $\mathcal{Q}_j(t) := \frac{\partial Q_{T_j}(t|d,x)}{\partial t}$ for $0 \le t \le t_1 - d_3$, and $\mathcal{H}_j(t) := \Lambda_j(t_1-d_j) + \Upsilon_{j,1}(t+\eta_j|t_1-d_j,x,3)$ and $\mathcal{Q}_j(t) := \frac{\partial Q_{T_1,T_j}(t|d,x)}{\partial t_1\partial t}$ for $t > t_1 - d_3$. Finally, $g_j{:=}\lambda_1(t_1)\phi_1(x)\phi_j(x)$ and we supress dependence of $\Lambda_1(t_1)$ and $\phi_j(x)$ on t_1 and x, respectively.

The equations (A.9), (A.10), by using the definitions of the previous paragraph, imply that we have the following system of two differential equations for almost all $t > 0$

$$\begin{aligned}&\frac{d}{dt}\mathcal{H}(t) = f\left(t, \mathcal{H}(t)\right),\\ &\quad\mathcal{H}(\tau) = \gamma_\tau, \text{ for some specific } \tau \in (0, t_1-d_3) \qquad \text{(initial conditions)},\end{aligned} \tag{A.11}$$

where $\mathcal{H} := (\mathcal{H}_2\ \mathcal{H}_3)'$ and $f := (f_2\ f_3)'$, with

$$f_j\left(t, \mathcal{H}\right) = \begin{cases}\Big[\mathcal{L}_G^{(2)}(\phi_1\Lambda_1(t), \phi_2\mathcal{H}_2, \phi_3\mathcal{H}_3)\phi_j\Big]^{-1}\mathcal{Q}_j(t) & \text{if } 0 < t \le t_1 - d_3,\\ \Big[\mathcal{L}_G^{(12)}(\phi_1\Lambda_1, \phi_2\mathcal{H}_2, \phi_3\mathcal{H}_3)g_j\Big]^{-1}\mathcal{Q}_j(t) & \text{if } t > t_1 - d_j.\end{cases}$$

It is straightforward to verify that all the requirements of Lemma 1 of Drepper and Effraimidis (2012) are satisfied. Hence, $\mathcal{H}_1$ and $\mathcal{H}_2$ are uniquely determined on $\mathbb{R}_+$ (using also the fact that $\mathcal{H}_1(0) = \mathcal{H}_2(0) = 0$). By definition, identification of $\Upsilon_{j,1}(t|t_1 - d_j, 3, x)$ follows for each $t > d_3 - d_2$ with t_1, x, d_2, and d_3 be fixed. Since $\Upsilon_{j,1}(t|t_1 - d_j, 3, x)$ is either cadlag or caglad with respect to $t_1 - d_j$, identification of $\Upsilon_{j,1}$ for the case $t_1 > d_3$ is obtained. By utilizing all the results of the previous paragraphs we derive identification of $\Upsilon_{j,1}$ for the cases $0 \leq T_1 \leq d_2$, $d_2 \leq T_1 \leq d_3$, and $T_1 > d_3$.

For the identification of the remaining interaction effect functions, we briefly discuss the necessary steps which are similar to the preceding paragraphs. Regarding the identification of $\Delta_{1,2}$ and $\Delta_{3,2}$, there are two possible scenarios: $i)$ $d_2 < T_2 \leq d_3 - d_2$, $ii)$ $T_2 > d_3 - d_2$. We first identify $\Delta_{1,2}$ and $\Delta_{3,2}$ for the case $i)$. In particular, we let $d \in \mathbb{D}_\infty$ and we identify $\Delta_{1,2}$. Based on this result, we can also directly identify $\Delta_{1,3}$ by considering $d \in \mathbb{D}$. To jointly identify $\Delta_{1,2}$ and $\Delta_{3,2}$ for the the case $ii)$, we let $d \in \mathbb{D}$ and by making use of Lemma 1, we achieve identification. Finally, to jointly identify $\Delta_{2,3}$ and $\Delta_{1,3}$, we let $d \in \mathbb{D}$ and working analogously to the previous paragraphs as well as utilizing Lemma 1, we get the desired result. ■

Identification of the set of functions $\{\Delta_{j,kl} : j, k, l = 1, 2, 3,\ \ k \neq j \neq l, k < l\}$. We will restrict our attention to $\Upsilon_{3,12}$; the arguments for the identification of the other combinations of j, k, l are similar and thus we will omit the proof for the corresponding combinations. Two scenarios are possible: $i)$ $T_1 \leq T_2 + d_2 < T_3 + d_3$ and $ii)$ $T_2 \leq T_1 + d_1 < T_3 + d_3$.

We will analyze the case $i)$ as the proof for the case $ii)$ is completely analogous. We can write for all $t > 0$, almost all $0 < t_1 < d_2$, almost all $t_2 \leq d_3 - d_2$, $d \in \mathbb{D}$, and $x \in \mathbb{X}$,

$$\begin{aligned}\frac{\partial^2 Q_{T_1,T_2,T_3}(t_1, t_2, t|x)}{\partial t_1 \partial t_2} = \mathcal{L}_G^{(12)}(&\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t_2|t_1 - d_2, 1, x),\\ &\phi_3(x)\Upsilon_{3,12}(t|t_1 - d_3, t_2 + d_2 - d_3, 2, x))\\ &\times \lambda_1(t_1)\phi_1(x)\lambda_2(t_2)\phi_2(x)\delta_{2,1}(t_2|t_1 - d_2, 1, x),\end{aligned} \tag{A.12}$$

Likewise, for all $t > 0$, almost all $0 < t_1 < d_2$, almost every $t_2 > d_3 - d_2$, $d \in \mathbb{D}$, and $x \in \mathbb{X}$,

$$\begin{aligned}\frac{\partial^2 Q_{T_1,T_2,T_3}(t_1, t_2, t|x)}{\partial t_1 \partial t_2} = \mathcal{L}_G^{(12)}(&\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t_2|t_1 - d_2, 1, x),\\ &\phi_3(x)\Upsilon_{3,12}(t|t_1 - d_3, t_2 + d_2 - d_3, 3, x))\\ &\times \lambda_1(t_1)\phi_1(x)\lambda_2(t_2)\phi_2(x)\delta_{2,1}(t_2|t_1 - d_2, 1, x).\end{aligned} \tag{A.13}$$

The left hand side of the above equation is observed from the data. By Propositions 1 and

2, all the quantities on the right-hand side are known except for $\Upsilon_{3,12}$. Given that $\mathcal{L}_G^{(23)}$ is strictly decreasing in its arguments, the identification of $\Upsilon_{3,12}$ follows by using also the fact that $\Upsilon_{3,12}(t|t_1 - d_3, t_2 + d_2 - d_3, \mathcal{N}_{12}, x)$ is either cadlag or caglad in $(t_1 - d_3, t_2 + d_2 - d_3)$. If $d_2 < t_1 < d_3$, the steps are almost identical by replacing $\phi_2(x)\Upsilon_{2,1}(t_2|t_1 - d_2, 1, x)$ with $\phi_2(x)\left(\Lambda_2(t_1 - d_2) + \Upsilon_{2,1}(t_2|t_1 - d_2, 2, x)\right)$ and $\delta_{2,1}(t_2|t_1 - d_2, 1, x)$ with $\delta_{2,1}(t_2|t_1 - d_2, 2, x)$. Similarly, if $t_1 > d_3$ we are encountered with a single subcase and we replace $\phi_2(x)\Upsilon_{2,1}(t_2|t_1 - d_2, 2, x)$ with $\phi_2(x)\left(\Lambda_2(t_1 - d_2) + \Upsilon_{2,1}(t_2|t_1 - d_2, 3, x)\right)$ and $\delta_{2,1}(t_2|t_1 - d_2, 1, x)$ with $\delta_{2,1}(t_2|t_1 - d_2, 3, x)$. ■

Proof of Proposition 2

The identification strategy we follow is the same as in the proof of Proposition 1. Note that, by construction, we always have $\mathcal{N}_k = \mathcal{N}_{kl} = 3$ and consequently, we will omit for notational simplicity this information.

Identification of the set of functions $\{\Lambda_j, \phi_j : j = 1, 2, 3\}$ and G. The result is directly obtained by making use of the distribution of $\{\min_{j\in\{1,2,3\}}(T_1, T_2, T_3), \arg\min_{j\in\{1,2,3\}}(T_1, T_2, T_3)\}|\{x\}$ and the identification result of Abbring and van den Berg (2003a). ■

Identification of the set of functions $\{\Delta_{j,k} : j, k = 1, 2, 3,\ j \neq k\}$. We will give in outline the proof of the joint identification of $\Upsilon_{2,1}$ and $\Upsilon_{3,1}$ which, by definition, uniquely determine the quantities $\Delta_{2,1}$ and $\Delta_{3,1}$, respectively. The (joint) identification of $\Upsilon_{1,2}, \Upsilon_{3,2}$ and also $\Upsilon_{1,3}, \Upsilon_{2,3}$ can be derived in a similar manner and as consequence, we will not discuss here these two cases.

Now, for any $x \in \mathbb{X}$ and almost all $t > 0$, we have

$$\lambda_j(t) = \left[\mathcal{L}_G^{(j)}(\phi_1(x)\Lambda_1(t), \phi_2(x)\Lambda_2(t), \phi_3(x)\Lambda_3(t))\phi_j(x)\right]^{-1} \frac{\partial Q_{T_j}(t|x)}{\partial t}. \tag{B.1}$$

Similarly, we obtain for each $x \in \mathbb{X}$, almost all $0 < t_1 < t$, and $j = 2, 3$

$$\lambda_j(t)\delta_{j,1}(t|t_1, x) = \Big[\mathcal{L}_G^{(1j)}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)\left(\Lambda_2(t_1) + \Upsilon_{2,1}(t|t_1, x)\right), \phi_3(x)(\Lambda_1(t_1) + \Upsilon_{3,1}(t|t_1, x)))\phi_j(x)\lambda_1(t_1)\phi_1(x)\Big]^{-1} \frac{\partial^2 Q_{T_1,T_j}(t_1, t|x)}{\partial t_1 \partial t}. \tag{B.2}$$

The equations (B.1) and (B.2) imply that we have a system of two differential equations. Following similar arguments to the proof of Proposition 1 and employing the result of Lemma of Drepper and Effraimidis (2012), we can solve with respect to $\Upsilon_{2,1}(t|t_1, x)$ and $\Upsilon_{3,1}(t|t_1, x)$. Using the fact that the latter quantities are either cadlag or caglad with respect to t_1, the identification of $\Upsilon_{2,1}$ and $\Upsilon_{3,1}$ follows. ■

Identification of the set of functions $\{\Delta_{j,kl} : j, k, l = 1, 2, 3,\ k \neq j \neq l, k < l\}$**.** We will restrict our attention on $\Upsilon_{3,12}$ which automatically, by definition, yields identification of $\Delta_{3,12}$. the arguments for identification of the other combinations of j, k, l are similar and thus we will omit the proof for these cases. There are two possible scenarios: $i)$ $T_1 < T_2 \leq T_3$ and $ii)$ $T_1 < T_3 \leq T_2$.

For all $t > 0$ and almost all $0 < t_1 < t_2 < t$, we have

$$\begin{aligned}\frac{\partial^2 Q_{T_1,T_2,T_3}(t_1, t_2, t|x)}{\partial t_1 \partial t_2} = \mathcal{L}_G^{(12)}(&\phi_1(x)\Lambda_1(t_1), \phi_2(x)\left(\Lambda_2(t_1) + \Upsilon_{2,1}(t_2|t_1, x)\right), \phi_3(x)\Lambda_3(t_3) \\ &\phi_3(x)\left(\Lambda_3(t_1) + \Upsilon_{3,1}(t_2|t_1, x) + \Upsilon_{3,12}(t|t_1, t_2, x)\right)) \\ &\times \lambda_1(t)\phi_1(x)\phi_2(x)\lambda_2(t_2)\delta_{2,1}(t_2|t_1, x).\end{aligned} \tag{B.3}$$

The left-hand side of the above equation is observed from the data. By the two previous results, all the quantities on the right-hand side are known except for $\Upsilon_{3,21}$. Given that $\mathcal{L}_G^{(23)}$ is strictly decreasing in its arguments, the identification of $\Upsilon_{3,12}$ follows (using also the fact that $\Upsilon_{3,12}(t|t_1, t_2, x)$ is either cadlag or caglad in (t_1, t_2) for any $t_1, t_2 > 0$ and $x \in \mathbb{X}$). Employing the statements of the two preceding results we prove the identification of $\Delta_{3,12}$ for the case $i)$ The steps are very similar for the case $ii)$and thus are omitted. The proof is complete. ■

Proof of Proposition 3

Proof of Proposition 3. It is straightforward, by Assumption B.2, to show that for all $t \in (t_a, t_b), \chi_1 \in \mathcal{P}_\chi$, and $d \in \mathbb{D}_{2\infty}$,

$$\mathbb{P}\left[T_1 > t | d, \chi_1\right] = \mathcal{L}_{G_1}(\mathcal{L}_{G_1}^{-1}(\tilde{\Lambda}(t, \chi_1)) + C) \tag{C.1}$$

for some $C \neq 0$. Applying the result of Brinch (2008), identification of $\tilde{\Lambda}$ follows. Next, for any $t > 0$, $\chi_1, \chi_2, \chi_3 \in \mathcal{P}_\chi$, and $d \in \mathbb{D}$,

$$\mathbf{P}\left[\bigcap_{j=1}^{3}(T_j + d_j > t + d_3)\middle| d, \chi_1, \chi_2, \chi_3\right] = \mathcal{L}_G(\tilde{\Lambda}(t + d_3, \chi_1), \tilde{\Lambda}(t + d_3 - d_2, \chi_2), \tilde{\Lambda}(t, \chi_3)) \tag{C.2}$$

By continuity of $\tilde{\Lambda}(., \chi)$ for any $\chi \in \mathbb{X}$ and varying appropriately t, d_2, and d_3, we identify $\mathcal{L}_G$ which yields identification of G. The identification methodology of the functions which capture the interaction effects is completely analogous to the proof of Proposition 1 and thus the details are omitted. ■

Proof of Proposition 4

Proof of Proposition 4. Consider the scenario $\zeta(t) = \zeta_1(t) = \zeta_2(t) = \zeta_3(t)$ for all $t > 0$, that is, all members in the group are characterized by the same realized covariate paths. Then, for $t > 0, \zeta \in \mathcal{P}_\zeta$,

$$\begin{aligned}\mathbb{P}\left[\bigcap_{j=1}^{3} T_j > t \middle| \zeta\right] &= \mathcal{L}_G(\tilde{\Lambda}(t, \zeta), \tilde{\Lambda}(t, \zeta), \tilde{\Lambda}(t, \zeta)) \\ &= \mathcal{L}_{\tilde{G}}(\tilde{\Lambda}(t, \zeta))\end{aligned}$$

with $\tilde{G}$ being the distribution of the random sum $V_1 + V_2 + V_3$. Applying the result of Brinch (2008), we achieve identification of $\tilde{\Lambda}$ and $\tilde{G}$. Next, we have for $t > 0$ and $\zeta_1, \zeta_2, \zeta_3 \in \mathcal{P}_\zeta$,

$$\mathbb{P}\left[\bigcap_{j=1}^{3} T_j > t \middle| \zeta_1, \zeta_2, \zeta_3\right] = \mathcal{L}_G(\tilde{\Lambda}(t, \zeta_1), \tilde{\Lambda}(t, \zeta_2), \tilde{\Lambda}(t, \zeta_3)).$$

By Assumption B*.5, the arguments of the Laplace Transform attain all values in an open subset of $\mathbb{R}_+^3$ which in turn, by the real analyticity property, yields identification of $\mathcal{L}_G$ and consequently of G. The identification strategy for the interaction effects is the same with the proof of Proposition 2 and therefore the details are omitted. ■

References

[1] Abbring, J. H. and G. J. van den Berg (2003a). The identifiability of the mixed proportional hazards competing risks model. *Journal of The Royal Statistical Society Series* B 65, 701-710.

[2] Abbring, J. H. and G. J. van den Berg (2003b). The nonparametric identification of treatment effects in duration models. *Econometrica* 71, 1491-1517.

[3] Abbring, J. H. and J. J. Heckman (2007). Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation. *Handbook of Econometrics*, Chapter 72 6 B, 5145-5303.

[4] Borjas, G. J.(1991). Ethnic Capital and Intergenerational Mobility. *The Quarterly Journal of Economics* 107, 123-150.

[5] Brinch, C. (2007). Nonparametric identification of the mixed hazards model with time-varying covariates. *Econometric Theory* 23, 349-354

[6] Blume, Lawrence E., William A. Brock, Steven N. Durlauf and Yannis M. Ioannides, 2010. Identification of Social Interactions. Economics Series 260, Institute for Advanced Studies.

[7] Van den Berg, G. J. (2001). Duration models: specification, identification and multiple durations. *Handbook of Econometrics*, Chapter 55 5, 3381-3460.

[8] Van den Berg, G. J., and B. Drepper (2011). A Unique Bond: Does Losing Your Co-Twin Affect your Remaining Life-Span? Working paper, Mannheim University.

[9] Brock, W. A. and Durlauf (2001). Interactions-based models. *Handbook of Econometrics*, Chapter 54 5, 3297-3380.

[10] Case, A. and L. F. Katz (1991). The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths. NBER Working Paper No. 3705.

[11] Cohen-Cole, E. and M. Fletcher (2008). Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics* 27, 1382-1387.

[12] Drepper, B. and G. Effraimidis, (2012). Nonparametric Identification of Dynamic Treatment Effects in Competing Risks Models. CDSE Working paper, Mannheim University.

[13] Elbers, C. and G. Ridder (1982). True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model. *The Review of Economic Studies*, 49, 403-409.

[14] Freund, John E. 1961. A Bivariate Extension of the Exponential Distribution. *Journal of the American Statistical Association* 56, No. 296, 971-977

[15] Gaviria, A. and S. Raphael (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83, 257-268.

[16] J. J. Heckman and B. E. Honoré (1989). The Identifiability of the Competing Risks Model. *Review of Economic Studies*, 77, 1138-1163.

[17] Honoré, B. E. and A. De Paula (2010). Interdependent Durations. *Review of Economic Studies*, 77, 1138-1163.

[18] Honoré, B. E. and A. De Paula (2011). Interdependent Durations in Joint Retirement, Working paper, Center for Retirement Research at Boston College.

[19] Kawaguchi D. (2004) Peer effects on substance use among American teenagers *Journal of Population Economics*, 17, 351-367.

[20] Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60, 531-542.

[21] Moffitt, R. A. (2001). Policy Interventions, Low-Level Equilibria, and Social Interactions. *Social Dynamics, Chapter 3*, 45-82.

[22] Monstad, K., C. Propper and Kj. G. Salvanes. (2011) Is Teenage Motherhood Contagious? Evidence from a Natural Experiment. Working Paper No. 11/262, Centre for Market and Public Organisation, Bristol Institute of Public Affairs.

[23] De Paula, A. (2009). Inference in a synchronization game with social interactions. *Journal of Econometrics*, 148, 56-71.

[24] Ridder, G., and T. Woutersen (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica*, 71, 1579-1589.

[25] Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, 116, 681-704.

[26] The NLSY79 (2005) *National Longitudinal Surveys Handbook, Chapter 3*, 34–52. *Econometrica*, 71, 1579-1589.

[27] Walter, W. (1998). *Ordinary Differential Equations*. Springer-Verlag. New York.