The Multivariate Mixed Proportional Hazard Model: Applications and Extensions

Inauguraldissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

Bettina Drepper

vorgelegt im Herbst 2013

Abteilungssprecher:	Prof. Dr. Eckhard Janeba
Referent:	Prof. Gerard van den Berg, Ph.D.
Korreferent:	Prof. Dr. Andrea Weber

Verteidigung: 8. Oktober 2013

Acknowledgements

First and foremost I would like to thank my supervisor Gerard van den Berg for his excellent guidance, continuing encouragement and the great research environment he created at his chair. He offered me the opportunity and freedom to pursue this research and supported my studies at the Center for Doctoral Studies in Economics at the University of Mannheim. I am also indebted to my co-advisor Prof. Dr. Andrea Weber for her encouragement and her constructive comments throughout my studies in Mannheim.
Furthermore, I would like to thank Bo Honoré and Aureo de Paula for their guidance during the end of my PhD studies and their support with my recent job search. The joint work with Georgios Effraimidis, co-author of chapter three, was and continues to be a great pleasure and source of motivation. Similarly, I greatly enjoyed the cooperation with Christian Arnold, a co-author of chapter four.
I am very grateful to my colleagues at the chair of Econometrics and Empirical Economics and my fellows from the Center for Doctoral Studies in Economics in Mannheim for creating a productive and enjoyable working environment and for giving constructive feedback in countless seminars. In particular I would like to thank Annette Bergemann, Petyo Bonev, Hans-Martin v. Gaudecker, Barbara Hoffmann, Lena Janys, Pia Pinger, Steffen Reinhold and Arne Uhlendorff.
Last but certainly not least, I thank Lukas Wiewiorra and my parents for their personal support, their permanent backing and encouragement.

Bettina Drepper

Mannheim, Fall 2013

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

This dissertation comprises four self-contained chapters that address questions from very diverse fields of research, including mortality research, social interactions, international cooperation, and statistical software development. While the empirical questions covered are interdisciplinary in nature and combine the field of economics with demography as well as political science, it is the underlying common methodology that connects all chapters. Specifically, each chapter addresses or uses a multiple duration framework that belongs to the class of multivariate mixed proportional hazard models or constitutes a variation or extension of this class of models.

The multivariate mixed proportional hazard model is a reduced-form multiple duration framework, with the marginal duration distributions each satisfying the popular mixed proportional hazard specification (see Van den Berg, 2001, for an overview). This model class is designed for use with clustered duration data, whereby spells of the same cluster are dependent due to unobservable determinants. The clustered spells often reflect several spells of the same unit, with examples including the life-spans of twin pairs (Chapter 2), the age of first substance use of siblings living in the same household (Chapter 3), or the time until several countries incorporate an international agreement into domestic law (Chapter 4). Here, the unobservable determinants causing a dependence within the cluster are shared family characteristics and the similar genetic makeup among siblings in the first two examples and unobserved characteristics of international agreements in the latter example.

With the timing-of-events approach, Abbring and van den Berg (2003b) introduce a second source of dependence to the multivariate mixed proportional hazard model for the case

of two parallel durations.[1] When duration 2 ends, the subsequent survival of duration 1 is affected via a treatment effect function. Numerous empirical studies use this model to evaluate the effects of treatments in a bivariate duration framework. While the main field of application is the evaluation of labor market programs (e.g., see Van den Berg et al., 2004; Lalive et al., 2005; Abbring et al., 2005), other examples include the effect of patent grants on the timing of licensing by start-up technology entrepreneurs (Gans et al., 2008), the effect of cannabis use on cocaine use (Van Ours, 2003) and school dropout (Van Ours and Williams, 2009), the effect of child birth on relationship duration (Svarer and Verner, 2008), and the effect of bereavement on the spouse's survival (Van den Berg et al., 2011). Variations and extensions to the timing-of-events model are presented in Chapters 2 and 3 with the aim of extending its usability to new fields of research. In Chapter 2, we introduce together with Gerard van den Berg a symmetric version of the timing-of-events approach to study bereavement effects in twin pairs. Furthermore, in Chapter 3 with Georgios Effraimidis, we investigate an extension to multiple spells to study social interaction effects between siblings for the first-time use of marijuana.

Chapters 4 and 5 address very diverse topics, with the only common theme being the underlying methodology. A multivariate mixed proportional hazard model is used in Chapter 4 to study inefficiencies in international cooperation in the Southern Common Market (with Christian Arnold and Gerard van den Berg). Finally, in Chapter 5, we consider together with Gerard van den Berg likelihood-based statistical inference with left-truncated data in the shared frailty model. This model constitutes a special case of the multivariate mixed proportional hazard model, whereby the unobserved determinants are shared within a cluster.

In Chapter 2, we address the question of whether losing your co-twin can affect your remaining life-span. Twins share a unique bond that can lead to severe emotional stress and health deterioration once, the bond is broken. We present new empirical evidence suggesting that the loss of the co-twin can shorten the remaining life-span of the surviving twin. The identification of such bereavement effects is severely complicated by the similar genetic makeup and early childhood experiences of twins, which constitute a major source of the dependence between twin life-spans and are typically not observed. Previous studies

[1] It should be pointed out that the timing-of-events approach is based on a censored data structure. Once duration 1 ends, this implies immediate right-censoring of duration 2.

of twin lifespans limit their attention exclusively to one of the two sources of dependence, namely the bereavement effect or genetic factors. We present a new identification result on a symmetric version of the timing-of-events model of Abbring and Van den Berg (2003), which enables us to unite both strands of the literature in one model by exploiting only weak covariate variation. Our empirical analysis is based on 9,268 twin pairs from the Danish Twin Registry, with the results suggesting that a male twin who has lost his identical co-twin when he reached the age of 75 will live on average 1.6 years shorter due to the experience of this loss. This bereavement effect is less severe for non-identical twins or if the loss is experienced at a higher age.

In Chapter 3, we use multivariate duration methods to study social interaction effects. The behavior of interest in many studies of social interactions is characterized by a transition time, with examples including the time a person purchases a new product, moves out of the neighborhood or starts smoking. We introduce a new strategy to identify social interaction effects from grouped transition data that exploits information in the timing of transitions. In particular, we account for two sources of dependence between the behavior of members of a peer group: (1) Once a group member starts to smoke, this directly affects the subsequent risk of the other group members starting to smoke ('social interaction effect'). Such social interaction effects may be highly flexible and differ across group members, covariates and over successive transitions in the group; and (2) Group members may have similar unobserved characteristics such as risk attitudes or tastes ('correlated effect').

This approach is based on the timing-of-events model by Abbring and van den Berg (2003b), which we extend to multiple parallel spells with varying entry dates and a highly flexible pattern of successive interactions between different group members. Moreover, we present an identification result for this model. The definition of social interactions in terms of a lagged effect in time enables overcoming the reflection problem (Manski, 1993) in the presence of correlated effects, without making use of an exclusion restriction. Additionally, given that the model accounts for selection effects, it enables the study of peer effects 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 unobservable characteristics.

This new approach is used to study social interaction effects in the first-time use of marijuana among siblings growing up together in American households, using data from the NLSY79. We find that once the oldest sibling in the household starts using marijuana,

this makes it more likely for his younger siblings to engage in the same habit. Conversely, the marijuana use decision of the younger siblings does not seem to have a contagious effect on the behavior of the other siblings.

Chapter 4 is concerned with inefficiencies arising in international cooperation. Compliance is a key concern in international cooperation, given that contracting is based on the reliability of all partners. Typically, free-riding is believed to be its main impediment in international relations. We present an institutional outlier that solves this issue, albeit only at the cost of other strategic inefficiencies. In the Southern Common Market (Mercosur), its four member states of Argentina, Brazil, Paraguay and Uruguay cannot benefit from the cooperation efforts of others, yet are rather inclined to make false promises. International agreements normally become legally binding once a state incorporates the international contract at the domestic level. However, in Mercosur, policies are only legally binding when all four countries have passed a provision. Given Mercosur's weak provisions for enforcement, actors prefer to reap benefits from contracting alone, knowing that they can easily prevent the policy from taking effect at a later stage. We offer a formal model that demonstrates the consistency of our argument.

Empirically, we substantiate our claims with the complete incorporation record of all Mercosur regulations between 1994 and 2008. A striking feature of Mercosur's ratification record is that half of the regulations introduced before 2004 have not entered into force after five years. We jointly model the ratification hazards of the four members, accounting for the change of the economic and political environment in the four member countries over time, as well as unobserved regulation-specific characteristics. For regulations with high media coverage and political relevance, we find that political actors are more likely to first make an initial public promise to their Mercosur partners, before subsequently prolonging ratification at the domestic level when public interest is low.

Finally, Chapter 5 addresses inference in parametric multivariate mixed proportional hazard models with shared unobserved determinants. With multiple-spell duration data, an unobserved shared component for groups of spells is often specified in the form of a shared frailty term. We consider random-effects likelihood-based statistical inference if the duration data are subject to left-truncation. Such inference with left-truncated data can be performed in the Stata software package for parametric and semi-parametric shared frailty models. We show that with left-truncated data, the commands ignore the weeding-

out process before the left-truncation points, affecting the distribution of unobserved determinants among group members in the data, namely among the group members who survive until their truncation points. We critically examine studies in the statistical literature on this issue as well as published empirical studies that use the commands. Simulations illustrate the size of the (asymptotic) bias and its dependence on the degree of truncation. We provide a Stata command file for the parametric case that maximizes the likelihood function that properly takes account of the interplay between truncation and dynamic selection.

Chapter 2

A Unique Bond: Does Losing your Co-twin affect your Remaining Life-Span?[1]

2.1 Introduction

The lives of twins are fundamentally intertwined. Twins share a unique bond, and the life of one twin may have a direct influence on the life of the other. Studies with bereaved twins document how the loss of the co-twin can cause severe psychological stress that can also lead to health deterioration (e.g., see Woodward, 1988; Segal et al., 1995; Segal and Ream, 1998). The aim of this paper is to identify the effect of bereavement experienced in adulthood on the residual life expectancy of the surviving twin. Endogeneity concerns have hampered the empirical analysis of this effect so far (see Hougaard et al., 1992a). As a main obstacle, twins usually share childhood experiences and have a similar or identical genetic makeup. Such factors are typically unobservable and strongly influence health outcomes throughout the whole life cycle of both twins. Overall, it is a challenging task to capture the complex dependence structure between twin life-spans.

We introduce a new binary survival model to the twin mortality literature that allows to study the effects of bereavement on the subsequent survival of the bereaved twin and at the same time accounts for unobservable childhood or genetic effects. The new model is

[1]This chapter is joint work with Gerard van den Berg. We thank the Danish Twin Registry and Kaare Christensen for kindly allowing us to use their data.

used to estimate bereavement effects in 9,270 adult twin pairs born in Denmark between 1873 and 1930. Our results suggest that a male twin, who has lost his identical co-twin when he reached the age of 75, will live on average 1.6 years shorter due to the experience of the loss. This effect is less severe for non-identical twins or if the loss is experienced at a higher age.

Previous studies are limited to modeling either the bereavement effect or the influence of genetic factors. Tomassini et al. (2001; 2002)[2] and Hougaard et al. (1992a) use survival models for each twin life-span and include the life-span of the co-twin as an exogenous time dependent covariate. Hougaard et al. point out the problem with this approach. Their estimated effect does not only capture the effect of bereavement but also captures the dependence due to shared genetic factors. The other strand of literature focuses on this latter type of dependence. The most elaborate approach is based on a bivariate frailty model that specifies a flexible dependence structure between the frailty terms (e.g. Hougaard et al., 1992a,b; Yashin and Iachine, 1995a; Wienke et al., 2001). Here, the frailty terms reflect the influence of unobservable factors such as childhood or genetic effects.

In her discussion of Hougaard et al.'s (1992a) paper, Flourney (1992) argues that a super-model is needed that accounts for both effect simultaneously: the bereavement effect and the influence of unobservable correlated factors. In this paper we present such a unifying model. We include the life-span of the co-twin as a time-dependent covariate in the model and account for the endogeneity of this variable by including it in the model as a second equation. With a flexible dependence structure between the two frailty terms we account for the influence of childhood and genetic effects. Related models are used in empirical econometric studies on the effect of labor market programs on unemployment durations. Here, we adapt the established timing of events model by Abbring and Van den Berg (2003b) to fit our symmetric setup of twins and show that with minimal covariate variation in the data the components of this new model including the bereavement effect can be identified from the observed joint distribution of twin life-spans.

In our application we use data on 2,808 monozygotic and 6,462 dizygotic twin pairs from the Danish Twin Registry. As our analysis exploits the timing of deaths, it is advantageous

[2]Tomassini et al. (2002; 2001) use a model in which they match each bereaved twin to two not bereaved twins based on zygosity, age, and sex and compare the two resulting hazard rates after the age when bereavement takes place. Note that this method also ignores the endogeneity of the time of bereavement caused by shared genetic factors.

to observe as many exits as possible in the data. The Danish Twin Registry is one of the oldest existing twin datasets and allows us to use cohorts from 1873 to 1930, ensuring that 80.8% of the twins have uncensored exits before our window of observation ends in 2004. The drawback of using such old cohorts is the limited information available on the twin pairs. The Danish Twin Registry is designed as a medical dataset, providing information on dates of exit on a daily basis but has very limited information on other characteristics of the twins. In particular, the observable characteristics available do not vary within same-sex twin pairs.[3] The identification result of Abbring and Van den Berg (2003b) relies on the assumption of sufficient covariate variation within the unit of interest, the twin pair in our case. Consequently, their identification result does not apply to our dataset. In Section 2.3 we present a new identification result for a symmetric[4] version of the timing-of-events model that does not rely on this assumption.[5]

Our semi-parametric identification result has a wider relevance for the empirical study of parallel systems and networks and for epidemiological research. The symmetric timing-of-events model describes a very general setting in which two parallel durations are connected due to both, observable characteristics and unobservable time-constant factors, and at the same time the first exit potentially affects the survival of the other. In the most extreme case, the complete symmetry of our model allows for the two durations to be indistinguishable in terms of observable characteristics. So even if the durations can not be indexed (or the index is uninformative) and the only observable covariates are characteristics of the pair, our identification result still applies. This result is relevant in cases of datasets with limited observable covariates, such as data of old cohorts or datasets in which the available covariates, creating the otherwise necessary variation within duration pairs, are potentially endogenous and therefore have to be excluded.

Our model allows an estimation of bereavement effects among twins while controlling for the influence of shared genetic factors. However, the symmetric timing-of-events model, we present for this purpose, can also be used to estimate the dependence between twin life-spans caused by shared genetic factors while controlling for a potential additional

[3]The major part of the dataset comprises same-sex twin pairs, since less effort was put into following up on different-sex twins in the Danish Twin Registry.

[4]In contrast to the original timing-of-events model, the model we use here allows for treatment in both directions. Before the first exit occurs, both life-spans can potentially affect each other.

[5]Note that in contrast to the model by Abbring and van den Berg, in the identification result presented here a multiplicative structure is imposed on the bereavement effect function.

causal dependence. There is an extensive field of research with the purpose to quantify the influence of genetic factors on mortality using data sets similar to ours (e.g., see Hougaard et al., 1992a,b; Yashin and Iachine, 1995a). Our model allows a comparison of these different approaches. In our empirical analysis, we show that the magnitude of the estimated effects change considerably when either the bereavement effect or the influence of genetic factors is ignored.

Many bereavement studies focus on conjugal bereavement (e.g., see Bowling, 1987; Lichtenstein et al., 1998; Lindeboom et al., 2002; Manor and Eisenbach, 2003; Van den Berg et al., 2011). These studies find convincing evidence that the loss of a spouse can severely affect mortality shortly after bereavement. However, it remains unclear whether the measured effect on mortality originates exclusively from emotional stress since the loss of the spouse also greatly affects the everyday life of the surviving partner. In contrast to spouses, most adult twins have separate families and support systems. This suggests that a causal dependence between twin life-spans should be in large part attributed to the effect of emotional bereavement.

After a brief literature review of the link between bereavement and mortality in Section 2.2, we introduce the symmetric timing-of-events model and address identification in Section 2.3. In Section 2.4, we shortly introduce the twin dataset from the Danish Twin Registry. For the purpose of our empirical analysis, we impose some additional structure on the general symmetric timing-of-events model in Section 2.5 using functional forms that are well established in the twin mortality literature. Subsequently, our estimation results are presented in Section 2.6. We sum up with a discussion of our results in Section 2.7 and a brief outlook in Section 2.8.

2.2 The link between bereavement and mortality

The twin studies by Segal et al. (1995, 2002); Segal and Ream (1998) and Woodward (1988) document how the loss of the co-twin can cause severe emotional stress. The grief intensity for an identical (monozygotic) twin is often higher than that for other relatives or spouses (see Segal and Bouchard, 1993; Segal et al., 1995). Besides feelings of despair, depersonalization (numbness, shock), rumination (preoccupation with the deceased) and loss of control, bereaved twins also show symptoms such as loss of appetite and vigor, as well as other physical symptoms (Segal and Blozis, 2002).

According to Selye's General Adaptation Syndrome (1936; 1955), psychological stress can be a major cause of disease, given that chronic stress causes long-term biochemical changes.

> *Every stress leaves an indelible scar, and the organism pays for its survival after a stressful situation by becoming a little older.* (Hans Selye, Smith, 1969)

Stress theories are specifically applied to bereavement (e.g., see Stroebe and Stroebe, 1987; Stroebe et al., 1993), which is considered a stressful life event with physical health consequences.[6]

The psychological manifestations of grief are generally well-documented, and twin studies have established the existence of a strong psychological reaction to the loss of the co-twin. Furthermore, the work by Selye (1936; 1955) and Stroebe et al. (1987; 1993) explain the direct link between the emotional stress of bereavement and health outcomes. However, no empirical study to date has clearly established a causal dependence between bereavement and mortality for twins.

2.3 Model and identification result

In the following, we introduce the new bivariate model for twin life-spans, which constitutes a symmetric version of the timing-of-events model by Abbring and van den Berg (2003b) (see also Abbring and Heckman, 2007). Following a continuous duration framework, at age $t \in [0, \infty)$ each twin faces a certain risk of dying, given that he has survived up to this point (mortality hazard). Since we are interested in measuring the causal effect of the end of one life-span on the subsequent residual life-span of the other (bereavement effect), we specify the mortality hazard of each twin $j = 1, 2$ conditional on the realization of the life-span of the co-twin T_k. Additionally, we condition on observable characteristics of the twin pair x and the realization of frailty terms V_j to account for childhood effects and the genetic makeup. The resulting mortality hazards of the two twins are almost symmetric, i.e. their functional forms only differ due to the possibly different realizations of V_1 and V_2.

[6] Furthermore, Sanders (1999) integrates Selye's well-established general theory of stress in her Integrative Theory of Bereavement. She points out that besides the familiar stages of grief, patients also show physiological changes and consequent vulnerability to illness after bereavement (see also Sanders, 1980)

Model 2.1. *The hazard rates of* $T_1|(T_2 = t_2, x, V_1)$ *and* $T_2|(T_1 = t_1, x, V_2)$ *are given by*

$$\theta(t|T_2 = t_2, x, V_1) = \lambda(t)\phi(x)\delta(t, t_2, x)^{I(t>t_2)}V_1$$
$$\theta(t|T_1 = t_1, x, V_2) = \lambda(t)\phi(x)\delta(t, t_1, x)^{I(t>t_1)}V_2,$$

where the vector of frailties $V = (V_1, V_2)'$ *is assumed to be drawn from the bivariate distribution* $G(v_1, v_2)$ *and the bereavement effect function is multiplicative in two of its arguments* $\delta(t, t_k, x) = \delta_a(t - t_k)\delta_b(t_k, x)$.

In Model 2.1, as long as both twins are alive, each twin j faces a mortality hazard of $\lambda(t)\phi(x)V_j$. Once the co-twin dies, the mortality hazard of the surviving twin is rescaled by $\delta_a(t - t_k)\delta_b(t_k, x)$, reflecting the bereavement effect. Here, the first multiplicative term δ_a describes the dependence of the bereavement effect on the time passed since the loss occurred, while δ_b accounts for the dependence on the age at the time of bereavement and the observable variables x. $I(t > t_k)$ denotes an indicator function that is 1 if the loss has occurred and 0 otherwise. The function $\lambda(t)$ captures the dependence of the mortality hazard on age and $\phi(x)$ holds the effect of the covariates.

Note that given the observed twin pair characteristics x, Model 2.1 allows for two sources of dependence between life-spans T_1 and T_2, the first of which is reflected in the joint distribution of V_1 and V_2. For instance, for monozyotic twin pairs, we would expect a high positive correlation between the two terms, due to the twins' identical genetic makeup. The second type of dependence is reflected in the bereavement effect function $\delta(t, t_k, x)$. Note that, conditional on x and V, the only dependence between life-spans T_1 and T_2 comes from the bereavement effect function $\delta(t, t_k, x)$. Consequently, this function can be given a causal interpretation as the effect of the end of one life-span on the other.

In contrast to the frailty terms V that reflect the influence of all time constant unobserved characteristics such as the genetic makeup, the bereavement effect accounts for the timing of deaths. This is why the bereavement effect in Model 2.1 can be seen as a local effect, given that it only affects the hazard rate of the surviving twin after the exit of the other has occurred. Accordingly, the influence of time-constant unobservable factors V can be seen as a global effect, since characteristics shaped during childhood and genetic dispositions influence the mortality hazard of the two twins over their whole life-span, i.e. $\forall t \in [0, \infty)$. This terminology provides an intuition for the identifiability of Model

2.1. Since the bereavement effect is assumed to be a local effect, whereas the influence of upbringing and the genetic makeup is assumed to be global (time-constant), both can be distinguished.

The structure imposed on the conditional mortality hazards in Model 2.1 assures the clear distinction between the local bereavement and the global effect of unobservable factors. Note that the local characteristic of the bereavement effect in Model 2.1 rules out anticipatory effects, particularly a scenario in which a twin anticipates the future date of death of his co-twin and is affected by this knowledge to the degree that his mortality hazard today is affected. Furthermore, the unobservable influences V are assumed to be purely time-constant (global), thus ruling out unobservable shocks that affect both twins, such as local epidemics or major events within the family.

In contrast to Model 2.1, the functions λ and ϕ in the timing-of-events model of Abbring and van den Berg (2003b) are allowed to differ across the two hazards and only the hazard of duration 1 can be directly affected by the exit of duration 2.[7] In their paper, the authors already point out that their identification results can be extended in a straightforward manner to a setting in which the full distribution of $(T_1, T_2)|x$ is observable and both durations can potentially be affected by the exit of the other, similar to our setup. However, a different identification strategy is needed to identify Model 2.1 in which λ and in particular ϕ is the same in both hazards. The latter implies that all covariates in the vector x enter both hazards with the same value and have the same effect. The difficulty arises from this complete symmetry in the covariate effects $\phi(x)$.

The result by Abbring and van den Berg (2003b) uses the fact that, the two durations are competing risks until the first exit occurs. Therefore, their proof exploits an identification result of the mixed proportional hazard competing risk model (Abbring and van den Berg, 2003a). Such a competing risks model requires variation of the covariate effects across the two hazards in order to trace out the bivariate frailty distribution $G(v_1, v_2)$. In particular, it must be assumed that $(\phi_1(x), \phi_2(x))$ can attain all values over a nonempty open set $\Phi \subset (0, \infty)^2$ when x is varied over its support $\mathcal{X}$.[8] Since it holds in our symmetric setup that $\phi_1(x) = \phi_2(x) = \phi(x)$, we cannot exploit this exogenous variation across the two

[7] The original model also does not need the assumption that the treatment effect can be separated into two multiplicative parts.

[8] If $\phi_j(x) = \mathrm{e}^{\beta_j^T x}$ then it would be sufficient that the vector x has two continuous covariates that affect the hazard rates of both risks but with different nonzero coefficients, and that are not perfectly collinear.

hazards in our model.[9]

Although the original model of Abbring and van den Berg is in some respects more flexible than Model 2.1, given that it allows for different baseline hazards and regression component functions across the two equations, it is also more restrictive in the sense that it relies on sufficient variation of the covariate effects across the two hazards. Therefore, the symmetric case of Model 2.1 is not covered by their result. The main difference in terms of the identification strategy is that while Abbring and Van den Berg's result exploits the results from the mixed proportional hazard competing risk model, our identification strategy exploits the symmetry of the model, enabling us to use of the identification results of the univariate mixed proportional hazard model by Elbers and Ridder (1982). By imposing a multiplicative structure on the treatment effect function, we are able to split the hazard rate into three multiplicative parts reflecting the dependence on time t, observables x, and unobservable influences V, which is characteristic for a mixed proportional hazard model. We exploit this structure at several steps throughout our proof.[10]

For the purpose of identification, we impose the following assumptions:

Assumption 2.1. *The vector x is k-dimensional with $1 \leq k < \infty$ and $\phi : \mathcal{X} \to U \subset (0, \infty)$. The set $\mathcal{X} \subset \mathbb{R}^k$ contains at least two values.*

Assumption 2.2. *$\delta_a : \mathbb{R}_+ \to (0, \infty)$ with $\lim_{s \downarrow 0} \delta_a(s) < \infty$ and for $\delta_b : [0, \infty) \times \mathcal{X} \to (0, \infty)$ it holds that $\nexists\, c \in (0, \infty)$ s.t. $\delta_b(0, x) = c\phi(x)^{-1}\ \forall x \in \mathcal{X}$.*

Assumption 2.3. *For the function $\lambda : [0, \infty) \to (0, \infty)$ it holds that for all $t \in (0, \infty)$ $\lim_{s \downarrow t} \lambda(s) < \infty$ and has integral $\Lambda(t) := \int_0^t \lambda(\tau)\, d\tau < \infty,\ \forall\ t \geq 0$ and further*

$$\tilde{\Lambda}(t, s) := \int_s^t \lambda(\tau)\delta_a(\tau - s)\, d\tau < \infty, \quad \forall\ \{(t, s) \in [0, \infty)^2 : t > s\}.$$

[9]For twins it is very unlikely that observable characteristics such as sex or cohort will affect twin 1 systematically different compared to twin 2. In our dataset, twins are indexed according to their order of births. The firstborn has index 1 and the second index 2. But this information is extremely unreliable especially for the older cohorts. Note further, that since we use cohorts from 1873 onwards, we only have a very limited set of covariates available in our analysis none of which vary within same-sex twin pairs. Therefore, we can not rely on sufficient exogenous variation within twin pairs.

[10]Note that the identification results presented in this section can be straightforwardly extended to the case where the bereavement effect function differs between the two durations. Thus, if the two spells can be distinguished in the data, it is possible to identify two separate bereavement effects $\delta_1(t, t_2, x)$ and $\delta_2(t, t_1, x)$. The first measures the effect of the exit of duration 1 on duration 2 and the other the effect of the exit of duration 2 on duration 1. However, in most applications including our twin model the causal effect is symmetric.

For some a priori chosen t_0, t_0^* *and* x_0, *it holds that* $\int_0^{t_0} \lambda(\tau)\, d\tau = 1$, $\int_0^{t_0^*} \lambda(\tau)\delta_a(\tau)\, d\tau = 1$ *and* $\phi(x_0) = 1$.

Assumption 2.4. *V is an $\mathbb{R}_+^2$-valued time-invariant random vector $(V_1, V_2)'$ and is drawn from distribution G, which does not depend on x, and has a finite positive mean. G is such that $P(V \in (0,\infty)^2) = 1$. Furthermore, for all $(t,x) \in (0,\infty) \times \mathcal{X}$ $\lim_{s \downarrow t} E(V_j | T_j \geq s, T_k = t, x) = E(V_j | T_j \geq t, T_k = t, x)$.*

Assumption 2.5. *$\exists$ an open set $\Psi \in (0,\infty)^2$ with $t_1 > t_2 \; \forall \; (t_1,t_2) \in \Psi$ s.t. at all points $(t_1,t_2) \in \Psi$ the function $\Delta(t_1,t_2,x) = \tilde{\Lambda}(t_1,t_2)\delta_b(t_2,x)$ is continuously differentiable with respect to t_2.*[11]

Note that for Assumption 2.1, a single dummy variable x that does not need to vary across the two hazards suffices, provided that it has an effect. In such a case, $\phi(x)$ takes on only two values on $\mathcal{X}$. The timing-of-events model of Abbring and van den Berg (2003b) usually requires two continuous variables with different effects to assure identification. For our model, the most limited case of covariate variation in form of a single dummy variable suffices.

Assumption 2.3 restricts the baseline hazard function to be continuous from the right for all $t \in (0,\infty)$. Note that this does not rule out the piecewise constant case or most functional forms. Furthermore, given that this property only has to hold for strictly positive values, functional forms with $\lim_{s \downarrow 0} \lambda(s) = \infty$ such as the Weibull function are not ruled out. However, the initial jump of the bereavement effect has to have a finite limit. Consequently, functional forms of δ_a with $\lim_{s \downarrow 0} \delta_a(s) = \infty$ are excluded.

Note that, in contrast to Abbring and van den Berg (2003b), we do not make the assumption of varying covariates between the two durations, but rather impose a multiplicative structure on the treatment effect. We also use slightly different regularity assumptions because our proof exploits identification results from the mixed proportional hazard model.

With Assumptions 2.1-2.5, we formulate the following two propositions:

Proposition 2.1. *If Assumptions 2.1-2.4 are satisfied, then the functions $\lambda, \phi, \delta_a, \delta_b$ from*

[11] Alternative assumption 5: The open set $\Psi \in (0,\infty)^2$ could also exist for $t_1 < t_2 \; \forall \; (t_1,t_2) \in \Psi$ s.t. at all points $(t_1,t_2) \in \Psi$ the function $\Delta(t_2,t_1,x)$ is continuously differentiable with respect to t_1.

Model 2.1 are non-parametrically identified (up to a scaling factor) from the distribution of $(T_1, T_2)|x$.

Note that G remains undetermined in Proposition 2.1. This leads to:

Proposition 2.2. *If Assumptions 2.1-2.5 are satisfied, then Model 2.1, which is characterized by the functions* $G, \lambda, \phi, \delta_a, \delta_b$, *is non-parametrically identified (up to a scaling factor) from the distribution of* $(T_1, T_2)|x$.

Proof of Proposition 2.1*. Identification of* λ *and* ϕ: Let $Z = \min\{T_1, T_2\}$ be the minimum of the two durations T_1 and T_2. The survival function of $Z|x$ is given as (see Appendix 2.A.1 for details)

$$S_Z(t|x) = \int_0^\infty \mathrm{e}^{-\Lambda(t)\phi(x)w} \, \mathrm{d}G_W(w), \text{ with } W = V_1 + V_2. \tag{2.1}$$

Note that, due to the symmetry of Model 2.1, the distribution of Z has a hazard rate of the mixed proportional form: $\theta_z(t|x, W) = \theta(t|T_2 \geq t, x, V_1) + \theta(t|T_1 \geq t, x, V_2) = \lambda(t)\phi(x)W$ with frailty $W = V_1 + V_2$ drawn from distribution G_W. The results by Elbers and Ridder (1982), (see also Lancaster, 1990; Van den Berg, 2001, for an overview)[12] on the identification of the mixed proportional hazard model imply that, under Assumptions 2.1-2.4, the model in Equation (2.1), characterized by the functions λ, ϕ and G_W, is identified up to a scaling factor (see Appendix 2.A.1 for details).

Identification of δ_a: The survival function of duration T_j given x and given that the exit of the other duration occurred at $T_k = 0$, can be expressed as follows

$$S(t|T_k = 0, x) = \int_0^\infty \mathrm{e}^{-\int_0^t \theta(\tau|T_k=0,x,V_j)\, d\tau} \, \mathrm{d}G_{V_j|T_k=0,x}(v_j),$$

with $\theta(t|T_k = 0, x, V_j) = \lambda(t)\phi(x)\delta_a(t-0)\delta_b(0,x)V_j$. Here, we make use of the subset $T_j|(T_k = 0, x)$ of the observable bivariate distribution $(T_1, T_2)|x$. Here, duration k exits at time $T_k = 0$, and therefore the hazard of the other duration is affected by bereavement over the full interval $(0, \infty)$. Due to the multiplicative structure of the bereavement effect function, the distribution of $T_j|(T_k = 0, x)$ has a hazard rate of the mixed proportional form: $\theta(t|T_k = 0, x, V_j) = \tilde{\lambda}(t)\tilde{\phi}(x)V_j$ with $\tilde{\lambda}(t) = \lambda(t)\delta_a(t)$ and $\tilde{\phi}(x) = \phi(x)\delta_b(0,x)$. Again, the results by Elbers and Ridder imply that, under Assumptions 2.1-2.4, the mixed

[12] See also Kortram et al. (1995a) for the case of only two possible values for $\phi(x)$.

proportional hazard model defined by $\{\tilde{\lambda}, \tilde{\phi}, G_{V_j|T_k=0,x}\}$ is identified up to a scaling factor; moreover, given that λ is known, this also identifies δ_a. Note that a key feature of the mixed proportional hazard model is the independence of observable variables x and unobservable frailties V. In Appendix 2.A.2, we show that under Assumptions 2.1-2.4, the conditional frailty distribution $G_{V_j|T_k=0,x}$ does not depend on x. Furthermore, Assumption 2.2 states that the functions $\phi(x)$ and $\delta_b(0,x)$ are not proportional, assuring that the function $\hat{\phi}(x) = \phi(x)\delta_b(0,x)$ generates sufficient exogenous variation.

Identification of δ_b: In the following, we exploit information on the jump of the hazard rate at the moment of bereavement

$$\begin{aligned}\frac{\lim_{s\downarrow t}\theta(s|T_k=t,x)}{\theta(t|T_k=t,x)} &= \frac{\phi(x)\delta_b(t,x)\lim_{s\downarrow t}\delta_a(s-t)\lambda(s)E(V_j|T_j\geq s,T_k=t,x)}{\phi(x)\lambda(t)E(V_j|T_j\geq t,T_k=t,x)} \\ &= \delta_b(t,x)\lim_{s\downarrow t}\delta_a(s-t)\frac{\lim_{s\downarrow t}\lambda(s)}{\lambda(t)}. \end{aligned} \tag{2.2}$$

Assumptions 2.2 and 2.3 assure the existence of $\lim_{s\downarrow t}\delta_a(s-t)$ and $\lim_{s\downarrow t}\lambda(s)$. Accordingly, the second equality directly follows from Assumption 2.4, stating that $\lim_{s\downarrow t}E(V_j|T_j \geq s, T_k = t, x) = E(V_j|T_j \geq t, T_k = t, x)$. Note, that the left hand side of Equation 2.2 is observable for all $(t,x) \in (0,\infty)\times\mathcal{X}$. Since $\lim_{s\downarrow t}\delta_a(s-t)$, $\lim_{s\downarrow t}\lambda(s)$ and $\lambda(t)$ are known from previous steps, we can trace out the function $\delta_b(t,x)$ over $(0,\infty)\times\mathcal{X}$.[13]

□

Proof of Proposition 2.2*. Identification of* G: Recall that the functions $\lambda, \phi, \delta_a, \delta_b$ in Model 2.1 are identified under Assumptions 2.1-2.2. The only function that remains undetermined is the bivariate frailty distribution G. For this last step, we exploit information of the observed density $f(t_1,t_2|x)$ for $t_1 > t_2$ (see Appendix 2.A.3)

$$f(t_1,t_2|x) = c(t_1,t_2,x)\partial^2_{s_1,s_2}\mathcal{L}_G\big(\phi(x)(\Lambda(t_2)+\Delta(t_1,t_2,x)),\phi(x)\Lambda(t_2)\big), \tag{2.3}$$

with $c(t_1,t_2,x) = \lambda(t_1)\lambda(t_2)\phi(x)^2\delta_a(t_1-t_2)\delta_b(t_2,x)$ and $\Delta(t_1,t_2,x) = \tilde{\Lambda}(t_1,t_2)\delta_b(t_2,x)$. Note that all functions on the right hand side of Equation 2.3 are identified, apart from the cross derivative of the bivariate Laplace transformation $\partial^2_{s_1,s_2}\mathcal{L}_G(s_1,s_2)$, with arguments $s_1 = \phi(x)(\Lambda(t_2)+\Delta(t_1,t_2,x))$ and $s_2 = \phi(x)\Lambda(t_2)$. The Laplace transformation $\mathcal{L}_G(s_1,s_2)$ is known to be a completely monotone function. This property implies that

[13]Here, $\delta_b(0,x)$ is already known from the last identification step.

its cross derivative $\partial^2_{s_1,s_2}\mathcal{L}_G(s_1,s_2)$ is also completely monotone (see Appendix 2.A.3). Given that completely monotone functions are real analytic, and that real analytic functions are uniquely determined by their values on a nonempty open set, the function $\partial^2_{s_1,s_2}\mathcal{L}_G(s_1,s_2)$ can be identified on its whole support $S = [0,\infty)^2$ if we know all its values on a nonempty open set. In Appendix 2.A.3, we show that under Assumption 2.5, the function $(\phi(x)(\Lambda(t_2)+\Delta(t_1,t_2,x)),\phi(x)\Lambda(t_2))$ attains all values on a nonempty open set $\Upsilon \subset (0,\infty)^2$ when t_1 and t_2 vary over $\Psi \subset (0,\infty)^2$ with $t_1 > t_2$.[14] This identifies $\partial^2_{s_1,s_2}\mathcal{L}_G(s_1,s_2)$ and the integral $\int_0^{s1}\int_0^{s2}\partial^2_{s_1,s_2}\mathcal{L}_G(u_1,u_2)\,\mathrm{d}u_1\mathrm{d}u_2$ subsequently gives us $\mathcal{L}_G$. Due to the uniqueness of the Laplace transformation, G is uniquely determined. □

2.4 The Danish Twin Registry

In our empirical analysis we use data from the Danish Twin Registry. The registry was first established in 1954 with the goal of following up on all same-sex twins born since 1873 and surviving as twins at least until the age of 6. However, there is some selectivity in the very early cohorts, with twins who died young less likely to be included in the sample. Furthermore, most of the information on characteristics is only available for twins who survived January 1, 1943. Therefore, we restrict attention to twin pairs still alive at this date. However, this is not a serious limitation given that we are particularly interested in the effects of bereavement at higher ages. We use cohorts from 1873 to 1930, assuring that we observe the exit of most twins prior to January 1, 2004, when our window of observation ends. While the registry contains some different-sex twin pairs, most effort was devoted to following up on same-sex and particularly monozygotic twin pairs. We refer to Skytthe et al. (2002) for detailed descriptions of the registry and the way in which it has been collected.

As a result, our sample includes 2,806 monozygotic and 6,462 dizygotic twin pairs, 1,219 of which are different sex twin pairs. All twins are born between 1873 and 1930 and both twins in all pairs survived at least until January 1, 1943. The birth and death dates and resulting individual lifetime durations are observed in days. Individuals still alive on January 1, 2004 or had emigrated have right-censored durations. Overall, the death

[14] Note, that if $t_1 < t_2 \;\forall\; (t_1,t_2) \in \Psi$ then the same reasoning can be applied to the function $(\phi(x)\Lambda(t_1),\phi(x)(\Lambda(t_1)+\Delta(t_2,t_1,x)))$, which then holds the arguments of $\partial^2_{s_1,s_2}\mathcal{L}_G$ in Equation (2.3) for the case $t_1 < t_2$.

date is observed for 80% of the individuals in our sample. We observe the death cause for 94.4% of this group, which is classified according to the International Classification of Diseases(ICD) system, versions 5-8, at the 3-digit level. These are grouped into 12 categories, with the following groups being of specific interest: 'cardiovascular' (32.42%, death due to cardiovascular malfunctions or diseases), 'apoplexy' (14.13%), 'cancer' (26.03%, death due to malignant neoplasms), 'suicide' (1.03%), 'accidents' (3.7%) and 'other' (including death due to tuberculosis, other infectious diseases, diseases of the respiratory, digestive or uro-genital system).

For each twin pair in our sample, we observe zygosity, sex, year of birth, season of birth and region of birth. Note that none of the available covariates vary within the twin pair, apart from sex. In the previous section, we showed that our model does not rely on this kind of variation. The information on zygosity is very accurate with a misclassification rate below 5% (see Holm, 1983; Lykken, 1978). In our analysis, we use an indicator for being born in Copenhagen to distinguish between rural and urban areas in Denmark. The additional distinctions between small towns and rural areas outside of Copenhagen proved to be uninformative in our empirical analysis.

Besides having one of the oldest existing twin datasets in the world, the country of Denmark is particularly suited for mortality studies using individual lifetime data over a long time interval. At the beginning of our window of observation in the 1870s, Denmark already had a quite well-established and comprehensive health care system compared to the rest of Europe. This is of particular importance for our purposes, given that a functioning health care system dampens economic shocks that twin pairs are exposed to over their whole life. Moreover, there were also no major epidemics in Denmark between 1873 and 2004. Recent studies have compared international mortality levels for 1918, finding that Denmark stands out as the country with the lowest levels of excess mortality for the 1918–/1919 worldwide influenza pandemic (see Canudas-Romo and Erlangsen, 2008; Ansart et al., 2009). Furthermore, Denmark remained neutral in both World Wars, and despite being occupied by Germany during the Second World War, casualties were negligible compared to the rest of Europe. In summary, lifetime data from Denmark from the 1870s to present provides a dataset that is little affected by economic or direct health shocks, compared to the rest of Europe.

2.5 Model of twin life-spans

In the following, we apply the symmetric timing-of-events model introduced in Section 2.3 to the dataset of life-spans from the Danish Twin Registry. For this purpose, we need to impose additional structure on the functions λ, ϕ, δ and G in (Model 2.1). In particular, the vector of frailties (V_1, V_2) for each twin pair is assumed to be drawn from a Cherian bivariate Gamma distribution[15], which is often used in lifetime models for twins (see Yashin and Iachine, 1995b; Wienke et al., 2001, 2002) and allows the interpretation of the individual frailty term as the sum of a shared twin pair-specific term $\tilde{V}_0$ and an individual-specific term $\tilde{V}_1$:

$$V_j = \tilde{V}_0 + \tilde{V}_j \quad for\ j \in 1, 2.$$

Here, each term $\tilde{V}_1, \tilde{V}_2$ and $\tilde{V}_0$ is independently drawn from a Gamma distribution. With this structure, the bivariate Gamma distribution of (V_1, V_2) has identical marginal distributions, reflecting the symmetry of life-spans within twin pairs. Their mean is normalized to one, and consequently the joint distribution of (V_1, V_2) can be fully described by two parameters: the variance σ^2 of V_j and correlation ρ of V_1 and V_2. The latter is computed as the ratio of the shared and total variation $\rho = \frac{Var(\tilde{V}_0)}{Var(\tilde{V}_0 + \tilde{V}_j)}$. Recall that our sample includes monozygotic (MZ) and dizygotic (DZ) twin pairs. Accordingly, we estimate separate parameters for both types of zygosity: $\sigma^2_{MZ},\ \rho_{MZ}$ and $\sigma^2_{DZ},\ \rho_{DZ}$.
In the following, we denote the two life-spans of each twin pair by the vector of random variables $(T_1\ T_2)$ and their realizations by $(t_1\ t_2)$. The twin life-spans follow a distribution given by the bivariate survival function $S(t_1, t_2|x) = P(T_1 > t_1, T_2 > t_2|x)$. For a small part of the twin pairs in our sample, at least one twin is still alive at the end of the observation window on January 1, 2004. Since the censoring points are determined by the cohort of the twin pair, this assures independent censoring in our data. Furthermore, the life-span of one twin is right-censored for a few twin pairs, while the co-twin is observed to live past this censoring point. Here, right-censoring may occur due to immigration, for instance, which implies that we do not observe the exact time of bereavement

[15]Given that we have substantial left-truncation in our dataset, the Gamma distribution would still be a justified approximation if the true underlying distribution were to differ (Abbring and van den Berg, 2007).

for the co-twin. The first exit could occur any time between the censoring point and ∞. Consequently, we have to integrate over the respective interval to account for the occurrence of all possible exit times.[16]

$$S(t_1, t_2|x) = \begin{cases} S^*(t_1, t_1|x) - \int_{t_2}^{t_1} S_{t_2}(t_1, \tau|x)\, d\tau & \text{, for } t_1 \geq t_2 \\ S^*(t_2, t_2|x) - \int_{t_1}^{t_2} S_{t_1}(\tau, t_2|x)\, d\tau & \text{, for } t_1 < t_2 \end{cases}$$

$$\begin{aligned} \textit{with} \quad S^*(t_1, t_2|x) = &(1 + \sigma^2\phi(x)[\Lambda(t_1) + \Lambda(t_2)])^{-\frac{\rho}{\sigma^2}} \\ &(1 + \sigma^2\phi(x)\Lambda(t_1))^{-\frac{(1-\rho)}{\sigma^2}} (1 + \sigma^2\phi(x)\Lambda(t_2))^{-\frac{(1-\rho)}{\sigma^2}} \end{aligned} \tag{2.4}$$

and partial derivatives $S_{t_j}(t_1, t_2|x) = \frac{\partial S(t_1,t_2|x)}{\partial t_j}$, for $(j = 1, 2)$.

As already mentioned in the data section, our dataset only includes twin pairs for which both twins were still alive on January 1, 1943. This truncation of data has to also be reflected in the likelihood function. For this purpose, we denote the age of twin j on January 1, 1943 by $t_{j,age43}$. This leads to the survival function conditional on both twins surviving January 1, 1943

$$S(t_1, t_2|T_1 > t_{1,age43}, T_2 > t_{2,age43}, x) = S(t_1, t_2|x)S(t_{1,age43}, t_{2,age43}|x)^{-1}$$

With this, we can derive the likelihood contribution of a twin pair

$$\begin{aligned} L(t_1, t_2, c_1, c_2|x) = [\ & c_1c_2S(t_1, t_2|x) - c_1(1 - c_2)S_{t_2}(t_1, t_2|x) \\ & -(1 - c_1)c_2S_{t_1}(t_1, t_2|x) + (1 - c_1)(1 - c_2)S_{t_1,t_2}(t_1, t_2|x)\ ] \\ & S(t_{1,age43}, t_{2,age43}|x)^{-1}. \end{aligned} \tag{2.5}$$

Here, c_1 and c_2 denote the censoring indicators for T_1 and T_2 and $S_{t_1,t_2}(t_1, t_2|x) = \frac{\partial^2 S(t_1,t_2|x)}{\partial t_1 \partial t_2}$. Note that due to the specific functional form of the Cherian bivariate Gamma distribution, the likelihood function has a closed form[17]. The functional forms of S, S_{t_1}, S_{t_2} and S_{t_1,t_2} and their derivation are presented in Appendix 2.A.4.

[16] The integrals $\int_{t_2}^{t_1} S_{t_2}(t_1, \tau|x)\, d\tau$ and $\int_{t_1}^{t_2} S_{t_1}(\tau, t_2|x)\, d\tau$ are approximated with numerical integration methods.

[17] The only exception are the integrals over the interval of all possible bereavement times for censored twin pairs.

For the purpose of our empirical analysis, we also impose additional structure on the functions $\phi, \lambda, \delta_a, \delta_b$ of Model 2.1. The logarithm of $\delta_a(t)$ is specified as piecewise constant with three time intervals after bereavement occurred: the first year after bereavement, the second to fourth year and after the fourth year. The function $ln(\delta_b(t_k, x)) = \delta_l^{age} + \beta_\delta^\intercal x$ models the dependence of the bereavement effect on the twin's age at the time t_k that she/he experiences the loss of the co-twin and the dependence on observable characteristics x such as sex and zygosity. Here, p indicates the age interval in which the loss occurred: ages below 65, 66 to 79 and above 80. The covariate effects enter the hazard as $e^{\beta^\intercal x}$, which is the standard choice in mixed proportional hazard models and the duration dependence function λ is assumed to follow a flexible version of the Gompertz function, i.e. $\lambda(t) = e^{\alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3}$.

We choose a flexible baseline hazard that includes the Gompertz function as a special case for $\alpha_2 = \alpha_3 = 0$, which is often used in mortality models and is known to give an acceptable fit. Specifying the correct functional form for the baseline hazard is particularly important for our analysis, given that we are measuring the impact of intermediate events in a lifetime. For instance, if the baseline hazard function were specified to be too restrictive in terms of the slope at higher ages, this lack of flexibility would be reflected in the causal bereavement effect, which in most cases occurs at higher ages. Note that we use a very wide range of cohorts 1873 to 1930 in our analysis, for which the aging process has evidently changed over time. In particular, the life expectancy at higher ages has drastically increased between 1873 and 1930 (see Gavrilov and Nosov, 1985). In order to account for this change in the shape of the duration dependence function, we estimate separate sets of parameters α_{c1}, α_{c2} and α_{c3} for three different cohort groups $c \in \{1, 2, 3\}$: 1873-1899, 1900-1915 and 1916-1930.

With this structure, we can express our model in terms of the logarithm of the hazard rates of twin 1 and 2 conditional on observable and unobservable variables x and V and the realization of the other duration

$$\begin{aligned}
\ln \theta(t|T_2 = t_2, x, V_1) &= \alpha_{c1} t + \alpha_{c2} t^2 + \alpha_{c3} t^3 + \beta' x + I(t > t_2)(\delta_q^t + \delta_p^{age} + x'\delta^x) + \ln(V_1) \\
\ln \theta(t|T_1 = t_1, x, V_2) &= \alpha_{c1} t + \alpha_{c2} t^2 + \alpha_{c3} t^3 + \beta' x + I(t > t_1)(\delta_q^t + \delta_p^{age} + x'\delta^x) + \ln(V_2). \quad (2.6)
\end{aligned}$$

Here, $\delta_q^t, \delta_p^{age}, \delta^x$ are parameters that model the effect of bereavement. The indicator for the three time intervals after bereavement is denoted by $q = 1, 2, 3$ and $p = 1, 2$ is the

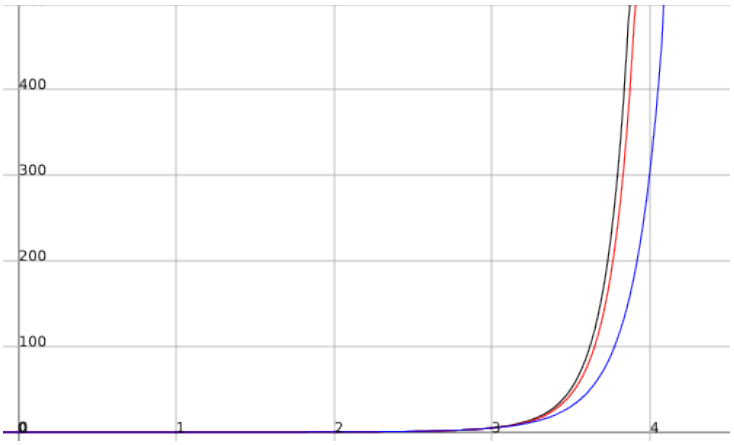
indicator for the three age groups at which bereavement occurs, with ages below 65 being the reference group.

Table 2.1: Estimation results of four bivariate survival models

	(Model I)		(Model II)		(Model III)		(Model IV)	
	Bivariate Model with BE		**Corr. Frailty Model no BE**		**Corr. Frailty Model with BE**		**Corr. Frailty Model with BE (ext.)**	
Variable	Estimate	St.Error	Estimate	St.Error	Estimate	St.Error	Estimate	St.Error
Covariates:								
male	.3982***	(.0167)	.6006***	(.0291)	.5318***	(.0299)	.513***	(.0301)
log(birth year)	-.1065***	(.0236)	-.2441***	(.0505)	-.1845***	(.0402)	-.1837***	(.0391)
spring	.0281	(.0186)	.0531*	(.0285)	.0428*	(.0248)	.0415*	(.0243)
Copenhagen	.0909***	(.0243)	.1313***	(.0371)	.1165***	(.0324)	.1144***	(.0318)
dizygotic	.1553***	(.0243)	.0472	(.0343)	.1103***	(.0352)	.1096***	(.0347)
Bereavement effect:								
first year	.5209***	(.0546)	-	-	.3838***	(.0779)	.4203***	(.09)
second to fourth year	.5224***	(.04)	-	-	.3816***	(.0714)	.409***	(.0844)
after four years	.5534***	(.0333)	-	-	.4037***	(.0774)	.3893***	(.0903)
dizygotic	-.2807***	(.0364)	-	-	-.24***	(.0752)	-.23***	(.0793)
male	-	-	-	-	-	-	.0353	(.0377)
age at ber. below 65	-	-	-	-	-	-	.1171***	(.0386)
age at ber. above 80	-	-	-	-	-	-	-.1253***	(.0503)
Corr. Gamma frailty:								
variance monozygotic	-	-	.6055***	(.1193)	.4059***	(.183)	.3894***	(.1957)
dizygotic	-	-	.5681***	(.1225)	.3474***	(.202)	.3424***	(.2129)
correlation monozygotic	-	-	.8676	(-)	.584	(-)	.5083	(-)
dizygotic	-	-	.4483	(-)	.3183	(-)	.1585	(-)

Note: The estimates represent the paramters β and $\delta_q, \delta_p, \delta^x$ in the likelihood function presented in Equation (2.6) using data on 9,268 twin pairs from the Danish Twin Registry (cohorts 1873 - 1930). Estimates with *, ** or *** reflect a 0.1, 0.05 or 0.01 significance level.

Figure 2.1: Flexible Gompertz-type baseline hazard functions by cohort groups



Note: The baseline hazard functions are based on the estimates of Model IV. A flexible version of the Gompertz baseline hazard is assumed: $\lambda(t) = \exp(\alpha_{c1}t + \alpha_{c2}t^2 + \alpha_{c3}t^3)$. Left line: cohort group 1873-1899, Middle line: cohort group 1900-1915, Right line: cohort group 1916-1930.

2.6 Empirical analysis

2.6.1 Estimation results

In our empirical analysis, we estimate four different bivariate survival models (Table 2.1: Models I-IV). Models III and IV are the models of interest, while Models I and II are used for the purpose of comparing our approach to previous models used in the twin mortality literature. Model II is a correlated frailty model that does not include a bereavement effect. It represents the strand in the epidemiological literature that models the influence of shared genetic factors by allowing for a dependence between frailty terms (see Yashin and Iachine, 1995a; Wienke et al., 2001). Note that a potential causal dependence between twin life-spans is ignored in these models. In Equations (2.4) - (2.6), this corresponds to the case of $\delta_a = \delta^{age} = \delta^x = 0$. On the other hand, Model I is a bivariate survival model, where the only dependence between twin life-spans conditional on covariates is modeled via a bereavement effect function. In fact, Model I does not allow for any influence due to unobservable characteristics of the twin pair. This corresponds to $\sigma^2 = 0$ in Equations (2.4) - (2.6) and represents the approach in the twin bereavement literature whereby bereavement is modeled as an exogenous event, ignoring the influence of shared genetic

factors (see Hougaard et al., 1992a; Tomassini et al., 2002). Finally, Model III represents the symmetric timing-of-events model that accounts for both the influence of shared genetic factors and a causal dependence between twin life-spans (Equations (2.4) - (2.6) with $\delta^{age} = \delta^{x} = 0$ [18]). In Model IV, we include a more flexible bereavement effect function allowing, besides zygosity, for a dependence on sex and the age at bereavement. This corresponds to the model in Equations (2.4) - (2.6).

When comparing the estimates of the correlated Gamma frailty distribution in Model II to those from Model III, one finds considerably higher estimates of the variance and the correlation parameters in Model II. This is true for the frailty distribution of monozygotic (σ^2_{MZ}, ρ_{MZ}), as well as that of dizygotic (σ^2_{DZ}, ρ_{DZ}) twin pairs. In particular, the correlation between frailties reflecting the influence of shared genetic factors decreases strongly (around 30%) when including the bereavement effect in Model III. It is clear from this comparison that the estimated correlation in Model II not only reflects the time-invariant influence of shared genetic factors but also captures some time-varying influences such as a causal dependence between twin life-spans.

In Model I, we find relatively high estimates for the bereavement effect, implying that a monozygotic male twin who is 75 years old and has lost his co-twin at the age of 70 would die on average 2.2 years earlier compared to if he had never experienced this loss. These high estimates are unsurprising given that they not only capture a bereavement effect but also the influence of shared genetic factors. We control for this influence in Models III and IV, finding considerably lower estimates (28% less in terms of residual life expectancy in Model IV). This illustrates how strongly the estimates of the bereavement effect are biased in the presence of unobserved shared genetic factors when the model fails to control for them. Considering these results, it also becomes clear to what extent previous empirical studies have overestimated a bereavement effect for twins.

Note that we do not report the estimated parameters of the baseline hazard function in Table 2.1. In Figure 2.1, the function $\lambda(t) = e^{\alpha_{c1}t+\alpha_{c2}t^2+\alpha_{c3}t^3}$ is plotted over the age interval 0-120 for the three cohort groups $c = 1, 2, 3$ implied by the estimated parameters in Model IV. Evidently, younger cohorts have a considerably lower mortality hazard at higher ages compared to the older cohorts. This change in the aging process over time is known as the late-life mortality deceleration (see Gavrilov and Nosov, 1985).

[18] The only exception to $\delta^{x} = 0$ is the dependence of the bereavement effect on zygosity.

In Model IV, we find a highly significant positive effect of being male (0.513), reflecting the shorter life expectancy for males compared to females. When comparing monozygotic male twins to monozygotic female twins, this estimate implies a higher residual life expectancy of 2.38 years for females at the age of 75 (see Tables 2a and 2b). Being born in spring has a weakly significant positive effect on the mortality hazard, in line with the findings of Doblhammer (2004). If a twin is born in Copenhagen, this increases mortality considerably (21.4% of the effect of being male), possibly due to a greater exposition to diseases, pollution or other risk factors in urban areas. Note that despite dizygotic twins being known to live slightly longer than monozygotic twins, we find a positive effect on mortality for dizygotic twins compared to monozygotic twins. However, considering that we restrict attention in this study to twins who survived infancy, this result may be explained by the fact that identical twins face a higher infant mortality risk compared to fraternal twins, leading to a selective sample that over-represents healthy identical twins.

We estimate a piecewise constant bereavement effect in Model IV, accounting for three different time intervals after bereavement occurred: the fist year after the loss, second to fourth year and after four years. The overall positive effect is highly significant and slightly decreases over time. Furthermore, the size of the bereavement effect strongly depends on zygosity, however we do not find evidence for a dependence on the sex of the twin. The size of the bereavement effect is almost twice the size for monozygotic compared to dizygotic twins. This large difference is in line with the findings from psychological studies (see Segal and Bouchard, 1993; Segal et al., 1995).[19] In Model IV, the bereavement effect function also depends on the age at bereavement. We distinguish between the ages before 65 and above 80, while ages 66 to 79 constitute the reference group. Evidently, there is a decrease of the effect of bereavement in the age at which the loss occurs. In particular, the effect of losing your co-twin after the age of 80 is relatively small, with an implied decrease in residual life expectancy of 0.58 years (for age 85, monozygotic males).

Table 2.2: Residual life expectancies for monozygotic male twins

Age	No Bereav.	Experienced Bereav. at age			
		60	70	80	90
65	14.22	11.22	-	-	-
75	7.85	5.82	6.26	-	-
85	3.31	2.27	2.48	2.73	-
95	.91	.58	.65	.72	.72

Note: Residual life expectancies in years for monozygotic male twins implied by the estimates of Model IV. The first column denotes the age of the twin. Columns 2 to 6 report the corresponding residual life expectancy, given that bereavement is never experienced, or experienced at ages 60, 70, 80 or 90.

Table 2.3: Residual life expectancies for monozygotic female twins

Age	No Bereav.	Experienced Bereav. at age			
		60	70	80	90
65	17.48	14.26	-	-	-
75	10.23	7.88	8.4	-	-
85	4.67	3.32	3.61	3.93	-
95	1.41	.92	1.01	1.13	1.13

Note: Residual life expectancies in years for monozygotic male twins implied by the estimates of Model IV. The first column denotes the age of the twin. Columns 2 to 6 report the corresponding residual life expectancy, given that bereavement is never experienced, or experienced at ages 60, 70, 80 or 90.

2.6.2 Residual life expectancies

One advantage of modeling twin life-spans at the individual level is the possibility to make predictions about residual life expectancies depending on the time when the loss is experienced. Expected residual lifetimes are relevant for health care policy and are frequently calculated within demographic and gerontological literature. The expected

[19]These studies conduct studies with bereaved twins and construct measures of grief intensities for monzygotic and dizygotic twins. Overall, they document grief intensities of monozygotic twins which are twice as large as the grief intensities observed for dizygotic twins.

Table 2.4: Residual life expectancies for dizygotic male twins

Age	No Bereav.	Experienced Bereav. at age			
		60	70	80	90
65	13.55	11.91	-	-	-
75	7.38	6.27	6.73	-	-
85	3.06	2.49	2.72	2.98	-
95	.83	.65	.72	.8	.8

Note: Residual life expectancies in years for monozygotic male twins implied by the estimates of Model IV. The first column denotes the age of the twin. Columns 2 to 6 report the corresponding residual life expectancy, given that bereavement is never experienced, or experienced at ages 60, 70, 80 or 90.

Table 2.5: Residual life expectancies for dizygotic female twins

Age	No Bereav.	Experienced Bereav. at age			
		60	70	80	90
65	16.77	15.01	-	-	-
75	9.7	8.41	8.95	-	-
85	4.35	3.62	3.92	4.26	-
95	1.29	1.02	1.12	1.25	1.25

Note: Residual life expectancies in years for monozygotic male twins implied by the estimates of Model IV. The first column denotes the age of the twin. Columns 2 to 6 report the corresponding residual life expectancy, given that bereavement is never experienced, or experienced at ages 60, 70, 80 or 90.

residual lifetime at age s is computed as follows (see Lancaster, 1990)

$$E(s) = \frac{\int_s^\infty S(t|x)\,dt}{S(s|x)}.$$

The residual life expectancies for male, female, monozygotic, and dizygotic twins implied by the estimates of Model IV (Table 2.1) are presented in Tables 2.2-2.5. A male monozygotic twin who has reached the age of 65 and lost his co-twin at the age of 60 will live on average for 11.22 remaining years. If he never had experienced this loss, he would live on average for 2 years longer (Tables 2.2) . A very similar pattern is observed

for female twins (Tables 2.3, 2.5). Since the dependence of the bereavement effect on sex is insignificant, we set this effect to zero in our calculations in Tables 2.2-2.5.

2.7 Discussion

The structure of the symmetric timing-of-events model in Equation (2.1) in Section 2.3 imposes some implicit assumptions on the underlying process generating the pairs of twin life-spans. Since identification of the model exploits the timing of the loss, a key assumption is that the event of losing your co-twin at age t does not affect your own mortality hazard prior to that date. Abbring and van den Berg (2003b) call this the 'no-anticipation' assumption. In the case of a severe long-term illness, a twin usually learns about the increased risk of dying of his co-twin when he is diagnosed with a severe illness before the actual loss occurs. However, this only constitutes a problem in terms of our model if the own mortality hazard reacts prior to the loss. In any case, some of the psychological symptoms of grief may already manifest at an early stage when the co-twin is diagnosed. We argue that following the true meaning of the word, 'bereavement' only takes place when the other person is really gone. Furthermore, the exact timing of the loss is usually not anticipated. Nevertheless, with long-term terminal illnesses the process of bereavement may to some degree already take place during the last stage of illness and the additional effect of loss will be small. In this case, our model would underestimated the true bereavement effect.[20] In light of this, one should interpret our estimated effect as the effect of actual bereavement, meaning the effect of physically losing the co-twin.

In the symmetric timing-of-events model all unobserved shared factors causing a dependence between the two life-spans of the twin pair are assumed to be time-invariant influences. In other words, our model accounts for all unobserved shared factors such as the genetic makeup or early childhood experiences as long as their influence on the mortality hazard is time-invariant. But some genetic dispositions manifest themselves more strongly during a certain stage in your life, leading to an increased mortality hazard. This additional source of dependence between twin life-spans would lead to an upward

[20]Consider the case in which a twin whose co-twin is diagnosed with a terminal illness is so severely affected by this anticipated loss that he himself dies before his co-twin. This very extreme case would constitute a problem for our model since anticipation would cause the estimated bereavement effect to capture a causal effect that is reverse.

biased bereavement effect. The investigation of this issue is left to further research. In particular, one could exploit the detailed information on death causes available in the Danish Twin Registry. In summary, it is conceivable that unobserved time-varying shared influences are partly responsible for the dependence between twin life-spans that our model can-not capture. However, in contrary to previous studies our model controls for the major source of dependence between twin life-spans, in the form of childhood and genetic makeup effects.

An additional source for unobserved time-varying shared variation are events that affect the health of both twins at the same time during their life. However, twins typically have their own family and support systems and often don't life in the same area. Furthermore, living in Denmark during the period 1873 to 2004 reduces the probability of being exposed to shocks on the national level such as major wars or epidemics. Additionally, the impact of health shocks is dampened by a well established health care system (see Section 2.4 for more details).

2.8 Conclusion

The contribution of this paper is twofold. First, we show that the symmetric version of the timing-of-events model (Model 2.1) can be identified from very limited covariate variation by imposing a multiplicative structure on the bereavement effect function. More specifically, the only exogenous variation that we exploit can be generated by a single dummy variable that does not need to vary between the two durations. The identification result of this symmetric model has wider relevance for the empirical study of parallel systems and networks and for epidemiological research. Model 2.1 can be applied to any symmetric bivariate duration model in which the dependence between durations is caused by two effects: the influence of time-invariant common factors and a causal effect. In particular, our identification result still applies if the two durations are not distinguishable from each other in any way, i.e. the index of duration 1 and 2 is completely uninformative and the only available covariates are characteristics of the duration pair.

Moreover, our empirical analysis unites two models that previously have exclusively been used separately by studies analyzing twin life-spans. The symmetric timing-of-events model allows to disentangle both effects of interest in this strand in the literature: the causal effect of bereavement and the influence of time-constant shared factors. Further,

our results reveal that previous studies that ignored the influence of childhood and genetic effects, severely overestimated the bereavement effect for twins.

2.A Appendix

2.A.1 Identification of λ and ϕ

The survival function of $Z|x$ with $Z = \min\{T_1, T_2\}$ is derived as follows

$$\begin{aligned} P(Z > t|x) &= P(T_1 > t, T_2 > t|x) \\ &= \int_0^\infty \int_0^\infty P(T_1 > t|x, V_1) P(T_2 > t|x, V_2) \, \mathrm{d}G(v_1, v_2) \\ &= \int_0^\infty \int_0^\infty \mathrm{e}^{-\phi(x)\Lambda(t)(V_1+V_2)} \, \mathrm{d}G(v_1, v_2) \\ &= \int_0^\infty \mathrm{e}^{-\phi(x)\Lambda(t)W} \, \mathrm{d}G_W(w) \qquad \textit{,with } W = V_1 + V_2. \end{aligned} \tag{2.7}$$

Note, that for the second equality we exploit that before the first exit occurs no bereavement effect will cause a dependence between T_1 and T_2. Consequently, conditional on x and V the events $(T_1 > t)$ and $(T_2 > t)$ are independent. We further use Assumption 4 which implies $G(v_1, v_2|x) = G(v_1, v_2)$.

In the following, we briefly discuss some of the assumptions used by Elbers and Ridder (1982) for the identification of a mixed proportional hazard model in view of the model given in Equation (2.7). Assumption 2.1 assures sufficient covariate variation in form of at least one dummy variable.[21] Further, the distribution of W has to be independent of x and has a positive and finite mean. Assumption 2.4 assures the independence of (V_1, V_2) and x. From this the independence of $W = V_1 + V_2$ directly follows. Similarly, as V_1 and V_2 are assumed to have finite positive mean, so does W.

2.A.2 Identification of δ_a

We consider the following hazard rate of mixed proportional form:

$$\theta(t|T_k = 0, x, V_j) = \tilde{\lambda_j}(t)\tilde{\phi_j}(x)V_j \;\; \textit{with } \tilde{\lambda_j}(t) = \lambda(t)\delta_a, \;\; \tilde{\phi_j}(x) = \phi(x)\delta_b(0, x), \tag{2.8}$$

[21] Also see Kortram et al. (1995b) for the case of only two possible values for $\phi(x)$.

where the frailties V_j are drawn from $G_{V_j|T_k=0,x}$ for $j,k \in \{1,2\}$ and $j \neq k$. One necessary assumption for the identifiability of this mixed proportional hazard model is that the frailty distribution does not depend on x. Note, that in the above model, the frailties V_j are drawn from a conditional distribution. Therefore, we need to show that $G_{V_j|T_k=0,x}$ does not depend on x. The conditional density of $V_j|(T_k = 0, x)$ is given by:

$$\begin{aligned} f(v_j|T_k = 0, x) &= \frac{\theta_k(0|x, V_j)S_k(0|x, V_j)f(v_j|x)}{\theta_k(0|x)S_k(0|x)} \\ &= \frac{\int_0^\infty \lambda(0)\phi(x)v_k \,\mathrm{d}G(v_k|x, V_j)f(v_j|x)}{\int_0^\infty \lambda(0)\phi(x)v_k \,\mathrm{d}G(v_k|x)} \\ &= \frac{E(V_k|x, V_j)f(v_j|x)}{E(V_k|x)}. \end{aligned} \tag{2.9}$$

According to Assumption 2.4 (V_1, V_2) are independent of x. Therefore, Equation (2.9) simplifies to

$$f(v_j|T_k = 0, x) = \frac{E(V_k|V_j)f(v_j)}{E(V_k)}. \tag{2.10}$$

From Equation (2.10) it also follows that the distribution of $(V_j|T_k = 0)$ for $j,k \in \{1,2\}$ and $j \neq k$ has a positive and finite mean, since $G(v_1, v_2)$ has this property.

2.A.3 Identification of G

Derivation of a mixing distribution: The density $f(t_1, t_2|x)$ for $t_1 > t_2$ can be expressed as follows

$$\begin{aligned} f(t_1, t_2|x) &= \int_0^\infty \int_0^\infty f(t_1|T_2 = t_2, x, V_1)f(t_2|x, V_2) \,\mathrm{d}G(v_1, v_2) \\ &= c(t_1, t_2, x) \int_0^\infty \int_0^\infty V_1V_2 e^{-\phi(x)(\Lambda(t_2)+\Delta_1(t_1,t_2,x))V_1} e^{-\phi(x)\Lambda(t_2)V_2} \,\mathrm{d}G(v_1, v_2) \\ &= c(t_1, t_2, x)\partial^2_{s_1,s_2}\mathcal{L}_G\big(\phi(x)(\Lambda(t_2) + \Delta_1(t_1, t_2, x)), \phi(x)\Lambda(t_2)\big), \end{aligned}$$

with $c(t_1, t_2, x) = \lambda(t_1)\lambda(t_2)\phi(x)^2\delta_a(t_1 - t_2)\delta_b(t_2, x)$, $\Delta(t_1, t_2, x) = \tilde{\Lambda}(t_1, t_2)\delta_b(t_2, x)$ and bivariate Laplace transformation $\mathcal{L}_G$ with cross derivative $\partial^2_{s_1,s_2}\mathcal{L}_G$.

Complete monotonicity: First, we state the definition of absolute monotonicity.

Definition 2.1. *Let Ω be a nonempty open set in $\mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}$ is absolutely monotone if it is nonnegative and has nonnegative continuous partial derivatives of all*

orders. f is completely monotone if $f \circ m$ is absolutely monotone, where $m : x \in \{\omega \in \mathbb{R}^n : -\omega \in \Omega\} \to -x$.[22]

Note, that this definition states that a function f is completely monotone if it's derivatives of all orders exist, and if these derivatives are continuous and have switching signs for each order (starting with a positive first derivative). From this definition it follows directly that if a function f is completely monotone then all derivatives of second order of f will also be completely monotone. Since the bivariate Laplace transformation $\mathcal{L}_G(s_1, s_2)$ is known to be a completely monotone function, it directly follows from Definition 1 that the cross derivative of $\mathcal{L}$ given by $\partial^2_{s_1,s_2}\mathcal{L}_G(s_1, s_2) = \frac{\partial^2 \mathcal{L}_G(s_1,s_2)}{\partial s_1 \partial s_2}$ is also completely monotone.

Tracing out the Laplace transformation: The function $f : \mathbb{R}^2_+ \to \mathbb{R}^2_+$ is given by $f(t_1, t_2) = (\phi(x)(\Lambda(t_2)+\Delta(t_1, t_2, x)), \phi(x)\Lambda(t_2))$. It maps the vector (t_1, t_2) on the vector of arguments of the Laplace transformation (s_1, s_2), with $s_1 = \phi(x)(\Lambda(t_2) + \Delta(t_1, t_2, x))$ and $s_2 = \phi(x)\Lambda(t_2)$. In the following we will show that we can vary (t_1, t_2) on an open set such that $f(t_1, t_2)$ will also attain all values in a nonempty open set. Under Assumption 2.5 (with $t_1 > t_2 \ \forall \ (t_1, t_2) \in \Psi$) it holds that at all points (t_1, t_2) in the open set Ψ the first derivatives of f exist and are continuous and f has Jacobian

$$J_f(t_1, t_2) = \begin{bmatrix} \phi(x)\lambda(t_1)\delta(t_1, t_2, x) & \phi(x)(\lambda(t_2) + \frac{\partial\Delta(t_1,t_2,x)}{t_2}) \\ 0 & \phi(x)\lambda(t_2) \end{bmatrix}.$$

Note, that the determinant of J_f is given by $det(J_f(t_1, t_2)) = \phi(x)^2\lambda(t_1)\lambda(t_2)\delta_1(t_1, t_2, x)$, and since under Assumptions 1-4 the functions $\phi, \lambda, \delta_a, \delta_b$ can only attain strictly positive (and finite) values on Ψ, it follows that $det(J_f(t_1, t_2)) \neq 0 \ \forall \ (t_1, t_2) \in \Psi$. Assumption 5 assures that $\frac{\partial\Delta(t_1,t_2,x)}{t_2}$ exists and is continuous on Ψ. Therefore, on the nonempty open set Ψ the function $f(t_1, t_2)$ is continuously differentiable with invertible Jacobian J_f. From the Inverse-Function Theorem it directly follows that there exists an nonempty open set $\Upsilon \subset (0, \infty)^2$ such that the function $f(t_1, t_2)$ attains all values in Υ when t_1 and t_2 vary over $\Psi \subset (0, \infty)^2$.

[22]For $n = 1$ this definition reduces to the familiar definitions in Widder (1946).

2.A.4 Derivation of the likelihood function

In the following the functional forms of S, S_{t_1}, S_{t_2} and S_{t_1,t_2} are derived. We start with the survival function $S(t_1, t_2|x) = P(T_1 > t_1, T_2 > t_2|x)$:

$$S(t_1, t_2|x) = \begin{cases} S^*(t_1, t_1|x) - \int_{t_2}^{t_1} S_{t_2}(t_1, \tau|x)\, d\tau & \text{, for } t_1 \geq t_2 \\ S^*(t_2, t_2|x) - \int_{t_1}^{t_2} S_{t_1}(\tau, t_2|x)\, d\tau & \text{, for } t_1 < t_2 \end{cases}$$

Recall, that in the case when the first exit is not observable due to censoring we have to integrate over all possible exit times. The resulting integrals $\int_{t_2}^{t_1} S_{t_2}(t_1, \tau|x)\, d\tau$ and $\int_{t_1}^{t_2} S_{t_1}(\tau, t_2|x)\, d\tau$ are approximated with numerical integration methods. Here, $S^*(t_1, t_2|x)$ denotes the survival function in the absence of a bereavement effect

$$\begin{aligned} S^*(t_1, t_2|x) &= \iint_0^\infty P(T_1 > t_1|x, V_1) P(T_2 > t_2|x, V_2)\, \mathrm{d}G(v_1, v_2) \\ &= \iiint_0^\infty \mathrm{e}^{\phi(x)\Lambda(t_1)(\tilde{V}_0+\tilde{V}_1)} \mathrm{e}^{\phi(x)\Lambda(t_2)(\tilde{V}_0+\tilde{V}_2)}\, \mathrm{d}G(\tilde{v}_0) dG(\tilde{v}_1) dG(\tilde{v}_2) \\ &= \int_0^\infty \mathrm{e}^{\phi(x)[\Lambda(t_1)+\Lambda(t_2)]\tilde{V}_0}\, \mathrm{d}G(\tilde{v}_0) \int_0^\infty \mathrm{e}^{\phi(x)\Lambda(t_1)\tilde{V}_1}\, \mathrm{d}G(\tilde{v}_1) \int_0^\infty \mathrm{e}^{\phi(x)\Lambda(t_2)\tilde{V}_2}\, \mathrm{d}G(\tilde{v}_2) \\ &= (1 + \sigma^2\phi(x)[\Lambda(t_1) + \Lambda(t_2)])^{-\frac{\rho}{\sigma^2}} (1 + \sigma^2\phi(x)\Lambda(t_1))^{-\frac{(1-\rho)}{\sigma^2}} (1 + \sigma^2\phi(x)\Lambda(t_2))^{-\frac{(1-\rho)}{\sigma^2}}. \end{aligned}$$

The last three equalities follow from the assumption that $G(v_1, v_2)$ is a Cherian bivariate Gamma distribution with independent terms $\tilde{V}_0, \tilde{V}_1, \tilde{V}_2$ drawn from Gamma distributions $\tilde{V}_0 \sim \Gamma(\rho\sigma^{-2}, \sigma^{-2})$ and $\tilde{V}_1, \tilde{V}_2 \sim \Gamma((1-\rho)\sigma^{-2}, \sigma^{-2})$.
In the following S_{t_j} is derived. For this purpose we define the functions g_a, g_b and g_c

$$\begin{aligned} g_a(s_1, s_2, x) &= 1 + \sigma^2\phi(x)[\Lambda(s_2) + \Delta(s_1|s_2, x)] \\ g_b(s_1, s_2, x) &= 1 + \sigma^2\phi(x)[2\Lambda(s_2) + \Delta(s_1|s_2, x)] \\ g_c(s, x) &= 1 + \sigma^2\phi(x)\Lambda(s). \end{aligned}$$

with $\Delta(s_1|s_2, x) = \int_{s_2}^{s_1} \lambda(u)\delta^t(u - s_2)\delta^{age,x}(s_2, x)\ \mathrm{d}u$.
We can now derive $S_{t_j}(t_j, t_k|x) = \frac{\partial S(t_j,t_k|x)}{\partial t_j} = -P(T_j = t_j, T_k > t_k|x)$. Let $t_j \geq t_k$ with

$j, k \in \{1, 2\}, j \neq k$

$$
\begin{aligned}
S_{t_k}(t_j, t_k|x) &= \iint_0^\infty P(T_j > t_j|T_k = t_k, x, V_j) P(T_k = t_k|x, V_k)\, \mathrm{d}G(v_j, v_k) \\
&= \phi(x)\lambda(t_k) \\
&\quad \iiint_0^\infty (\tilde{V}_0 + \tilde{V}_k) \mathrm{e}^{\phi(x)[\Lambda(t_k)+\Delta(t_j|t_k,x)](\tilde{V}_0+\tilde{V}_j)} \mathrm{e}^{\phi(x)\Lambda(t_k)(\tilde{V}_0+\tilde{V}_k)}\, \mathrm{d}G(\tilde{v}_0) dG(\tilde{v}_j) dG(\tilde{v}_k) \\
&= \phi(x)\lambda(t_k) g_b(t_j, t_k, x)^{-(\frac{\rho}{\sigma^2}+1)} g_c(t_k, x)^{-(\frac{(1-\rho)}{\sigma^2})} g_a(t_j, t_k, x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} \\
&\quad [\rho g_a(t_j, t_k, x) + (1-\rho) g_b(t_j, t_k, x)].
\end{aligned}
$$

This yields

$$
S_{t_j}(t_j, t_k|x) = \begin{cases}
\frac{\partial S^*(t_j, t_j|x)}{\partial t_j} + \int_{t_k}^{t_j} S_{t_1,t_2}(t_1, \tau|x)\, d\tau & , \text{for } t_j > t_k \\
\phi(x)\lambda(t_k) g_b(t_j, t_k, x)^{-(\frac{\rho}{\sigma^2}+1)} g_c(t_k, x)^{-(\frac{(1-\rho)}{\sigma^2})} & \\
g_a(t_j, t_k, x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} [\rho g_a(t_j, t_k, x) + (1-\rho) g_b(t_j, t_k, x)] & , \text{for } t_j \leq t_k.
\end{cases}
$$

Finally, $S_{t_1,t_2}(t_1, t_2|x) = \frac{\partial^2 S(t_1,t_2|x)}{\partial t_1 \partial t_2} = P(T_1 = t_1, T_2 = t_2|x) = f^*(\max\{t_1, t_2\}, \min\{t_1, t_2\})$ with

$$
\begin{aligned}
f^*(t_j, t_k) = \phi(x)^2 \lambda(t_j)\lambda(t_k) &\delta^t(t_j - t_k) \delta^{age,x}(t_k, x) \\
& g_b(t_j, t_k, x)^{-(\frac{\rho}{\sigma^2}+2)} g_a(t_j, t_k, x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} g_c(t_k, x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} \\
& [\rho(\rho + \sigma^2) g_a(t_j, t_k, x) g_c(t_k, x) + \rho(1-\rho) g_b(t_j, t_k, x) g_c(t_k, x) \\
& \rho(1-\rho) g_b(t_j, t_k, x) g_a(t_j, t_k, x) + (1-\rho)^2 g_b(t_j, t_k, x)^2].
\end{aligned}
$$

Chapter 3

Social Interaction Effects in Duration Models: The First-Time Use of Marijuana among Siblings[1]

3.1 Introduction

The study of social interactions has been of constant interest in economics and sociology over the past two decades (e.g., see Borjas, 1995; Manski, 2000; Brock and Durlauf, 2001a,b; Moffitt, 2001; Sampson et al., 2002; Calvó-Armengol et al., 2009) 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. [2] In this paper, we introduce a new strategy to identify social interaction effects from grouped transition data, using a multivariate duration framework. We use this model to study marijuana use of siblings, allowing for two sources of dependence between the behavior of siblings conditional on observable characteristics: (1) Once a teenager starts using marijuana, this may directly affect the subsequent risk of his siblings to engage in the same behavior ('social interaction effect'); (2) Siblings may behave similarly due to unobserved family characteristics ('correlated effect'). The definition of social interactions in terms of a

[1] This chapter is joint work with Georgios Effraimidis.

[2] Different versions of Manski's model are widely used in applications estimating peer effects (e.g., see Gaviria and Raphael, 2001; Sacerdote, 2001; Cohen-Cole and Fletcher, 2008).

lagged[3] effect in time allows overcoming the reflection problem in the presence of correlated effects without making use of an instrument as used by Case and Katz (1991) or Monstad et al. (2011) or an exclusion restriction as suggested by Moffitt (2001). Furthermore, given that our approach accounts for unobserved group characteristics, this 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 unobservable characteristics. Additionally, social interaction effects are highly flexible in 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 by siblings growing up together in American households[4]. Substance use is considered a highly social behavior (see Gaviria and Raphael, 2001; Kawaguchi, 2004). When a teenager uses marijuana for the first time, this may directly affect the subsequent behavior of his siblings through several different channels. To begin with, this transition may cause his siblings to copy his behavior. Alternatively, such a change in behavior within the household may reduce the stigma attached to using cannabis, 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 substance use, an effect of accessibility. One advantage of our approach is that it can distinguish[5] between the effect of the first transition and the effects of subsequent transitions within a group. Based on this distinction, to some extend accessibility and information effects may be separated from classical peer effects.[6]

[3]Given that our identification results are based on a model in continuous time, the period between transition and response is practically zero.

[4]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 effects.

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

[6]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.

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 chapter, we show that the joined observation of transition times allows to identify additional dynamics within 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 social status or 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.[7] 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 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.[8] This also captures the extreme case where no transitions have any effect, apart from the first. With marijuana 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 cannabis 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 initially targeted by such a policy. The distinct role of such 'key', 'high status', or 'influential' individuals in social networks is addressed by Ballester et al. (2006), Iyengar et al. (2009) and Aral and Walker (2012).

[7] 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.

[8] 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.

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 may be 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.[9] Varying entry points play a crucial role in many applications, because they determine the different starting points of an underlying risk process faced by each individual in the sample. In the case of siblings, this process represents the dependence of the risk to start using marijuana on age. Besides varying entry points, 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)[10] hazard type model that uses the information in the timing of transitions within groups to identify social interaction effects in the presence of correlated unobserved characteristics. The idea of exploiting the timing of events to disentangle a causal effect from a selection effect is introduced by Abbring and van den Berg (2003b), in the context of treatment evaluation in a bivariate duration framework. A symmetric version of this two spell model[11] is presented in Chapter 2 to study bereavement effects within twin pairs. Extending the timing-of-events approach to a general model of multiple parallel spells raises several new issues that are not encountered in the 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

[9] 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.

[10] In Section 3.2.3, we discuss conditions under which the proportionality assumption can be dropped, leading to a multivariate mixed hazard type model.

[11] 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 (also see Abbring and Heckman, 2007).

of events.[12] 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.[13]

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 of agents (see Brock and Durlauf, 2001b; Blume et al., 2010, for overviews). 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 of literature that uses equilibrium models with interdependent utility functions, we do not specify the underlying behavioral model of social interactions. Rather than assuming that the observed behavior represents an equilibrium outcome, we understand social interactions as a dynamic process of successive actions and reactions within a group. A key feature of the extended timing-of-events approach is that the transition hazard of a group member may directly react in response to transitions of other members.[14]

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. This corresponds to the 'no-anticipation' assumption of Abbring and van den Berg (2003b). This assumption states that individuals either do not anticipate the action of fellow peers, or if they do, they do not react to this anticipated action before it takes place. In applications where forward looking and strategic incentives dominate the behavior of group members, equilibrium models are more suitable to capture such dynamics (e.g., see Honoré and de Paula, 2010, 2013). In contrast, our approach focuses on applications where a transition of a group member is comparable to an unanticipated

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

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

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

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 marijuana at his school, the change in his behavior is similar to an unanticipated shock to his siblings.

Social interaction effects in cannabis use are studied by e.g. Kandel (1978); Kawaguchi (2004); Eisenberg (2004) and Clark and Loheac (2007) for American adolescents and by Pudney (2002); Moriarty et al. (2012) and McVicar and Polanski (2013) for adolescents living in the United Kingdom. Popular choices of reference/peer groups are the school grade, classmates or nominated friends.

Additionally, peer effects in smoking or substance use among siblings are addressed by several studies. Using a large dataset on cigarette smoking from the U.S. Current Population Surveys, Harris and González López-Valcárcel (2008) find that each additional smoking sibling in the household raised a young person's probability of smoking by 7.6%.[15] Krauth (2005) uses data on smoking among Canadian youths and finds a small positive sibling effect after controlling for selection effects, proxied by the degree of selection on observable variables. Similarly to our empirical study, Altonji et al. (2010) study substance use among siblings using data from the NLSY97. In contrast to our identification strategy, their approach relies on an exclusion restriction, i.e. it is assumed that only the substance use behavior of the older siblings can affect the behavior of the younger siblings but not vice versa, thereby circumventing the reflection problem. Based on this assumption, the authors find that marijuana use of younger siblings is positively affected by the example of their older siblings.

In our empirical analysis, we use data from the National Longitudinal Surveys (NLSY79), 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.

[15] Conversely, Chen (2010) finds, using sibling data from the NLSY79, a negative sibling effect for smoking cigarettes, implying a differentiation effect among siblings.

In the next section we introduce the multivariate duration model with social interaction effects and present our identification results. In Section 3.3 we discuss the data set, estimation method and results of our application. We conclude in Section 3.4.

3.2 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, given that all interesting dynamics occur within this setting. The extension to more than three spells is straightforward and will not be further discussed.

3.2.1 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 achieve 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-dimensional 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

$$\theta_j(t|T_k, T_l, d, x, V_j) = \begin{cases} \lambda_{j,0}(t, d, x)V_j & \text{if } t \leq \min\{T_{jk}, T_{jl}\}, \\ \lambda_{j,k}(t|T_{jk}, d, x)V_j & \text{if } T_{jk} < t \leq T_{jl}, \\ \lambda_{j,l}(t|T_{jl}, d, x)V_j & \text{if } T_{jl} < t \leq T_{jk}, \\ \lambda_{j,kl}(t|T_{jk}, T_{jl}, d, x)V_j & \text{if } \max\{T_{jk}, T_{jl}\} < t. \end{cases} \tag{3.1}$$

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 (3.1)

$$\delta_{j,k}(t|T_{jk}, d, x) := \frac{\lambda_{j,k}(t|T_{jk}, d, x)}{\lambda_{j,0}(t, d, x)} \tag{3.2}$$

$$\delta_{j,kl}(t|T_{jk}, T_{jl}, d, x) := \frac{\lambda_{j,kl}(t|T_{jk}, T_{jl}, d, x)}{\lambda_{j,q}(t|T_{jq}, d, x)} \quad \text{for } q = \arg\min_{k,l}\{T_{jk}, T_{jl}\} \ , \tag{3.3}$$

with (3.2) representing the effect of the exit of member k on the hazard of member j and (3.3) 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 (3.2) and (3.3). The functions $\lambda_{j,k}$ and $\lambda_{j,kl}$ are components of the conditional hazard rates $\theta_j(t|T_k, T_l, d, x, V_j)$ and are therefore not directly observable from the data. This poses an identification problem for the social interaction effect 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 (3.1). Firstly, the unobservable influences $(V_1\ V_2\ V_3)'$, which 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 (see Lancaster, 1992; Van den Berg, 2001, for an overview). Secondly,

in model (3.1), the effect of a transition of a member k only enters the hazard rate of member j after it occurs (for all $t > T_{jk}$). Known as the 'no-anticipation' assumption (see Abbring and van den Berg, 2003b), this restricts the dependence structure between the three transitions T_1, T_2 and T_3. It plays a crucial role for the identification and estimation of model (3.2) as it enables expressing the joint distribution of $\{T_1, T_2, T_3\}|\{d, x, V\}$ in terms of conditional distributions $\{T_j\}|\{T_k, T_l, d, x, V_j\}$.

In this section, we discuss different sets of assumptions under which the interaction effect functions (3.2) and (3.3) in Model (3.1) can be identified. We first consider the case of proportionality of the covariate effects leading to the popular mixed proportional hazard specification.

Model 3.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

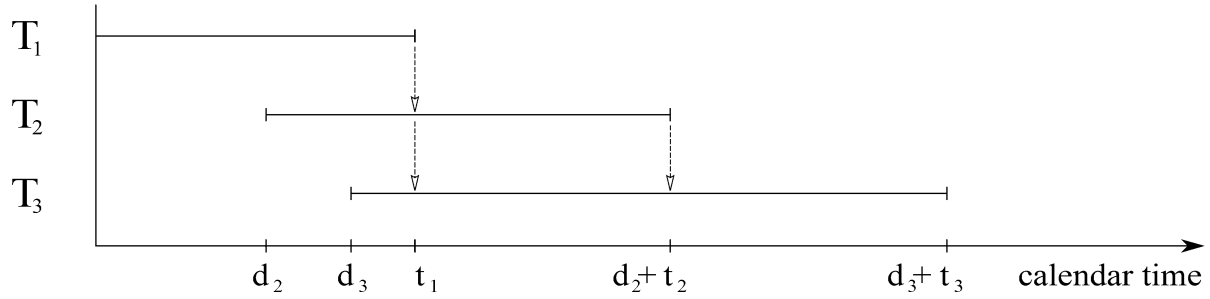
$$\delta_j(t|T_k, T_l, d, x) := \delta_{j,k}(t|T_{jk}, \mathcal{N}_k, x)^{I_{j,k}(t)}\delta_{j,l}(t|T_{jl}, \mathcal{N}_l, x)^{I_{j,l}(t)}\delta_{j,kl}(t|T_{jk}, T_{jl}, \mathcal{N}_{kl}, x)^{I_{j,kl}(t)},$$

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. 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 $\max\{d_k + T_k, d_l + T_l\}$, namely 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.

In order to provide some intuition for Model 3.A, we consider a concrete example in Figure 3.1. Here, the individual labeled as 1 (i.e. the individual who enters the risk process

Figure 3.1: Example of realized transition times



Note: Example for 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. The arrows represent the social interaction effects δ_j, affecting the subsequent transition hazards of the other two group members once member j transitions.

first) transitions into the state of interest (starts using 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, the 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 each member j is given by $\lambda_j(t)\phi_j(x)V_j$. 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. In this example, member 2 is the second to transition at duration $T_2 = t_2$, 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 3.A, we employ a set of assumptions formalized below.

Assumption 3.A.1. *The function $\phi_j : \mathbb{X} \to (0, \infty)$ is such that it attains all values on 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 3.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 3.A.3. *The G is does not depend on x and d. Moreover, for $j = 1, 2, 3$, $\mathbb{E}(V_j) < \infty$.*

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

$$\begin{aligned}
\Upsilon_{j,k}(t|s,\mathcal{N}_k,x) &:= \int_{\max\{0,s\}}^{t} \lambda_j(\omega)\delta_{j,k}(\omega|s,\mathcal{N}_k,x)d\omega, \\
\Delta_{j,k}(t|s,\mathcal{N}_k,x) &:= \int_{0}^{t} \delta_{j,k}(\omega|s,\mathcal{N}_k,x)d\omega, \\
\Upsilon_{j,kl}(t|s,y,\mathcal{N}_k,x) &:= \int_{\pi(s,y)}^{t} \lambda_j(\omega)\delta_{j,q}(\omega|\min\{s,y\},\mathcal{N}_q,x)\delta_{j,kl}(\omega|s,y,\mathcal{N}_{kl},x)d\omega, \\
\text{and } \Delta_{j,kl}(t|s,y,\mathcal{N}_{kl},x) &:= \int_{0}^{t} \delta_{j,kl}(\omega|s,y,\mathcal{N}_{kl},x)d\omega
\end{aligned}$$

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 3.A.1 states that there has to be sufficient variation of the covariate effects for each member. A sufficient condition for this assumption is the existence of a continuous group-level characteristic and continuity of the function ϕ_j. Moreover, it also imposes some innocuous normalization. Assumption 3.A.2 is not restrictive, given that it allows for several parametric choices for the baseline hazard. Additionally, it normalizes the integrated baseline hazard for some particular value. Assumption 3.A.3 is common in the analysis of the mixed proportional hazard model (see Elbers and Ridder, 1982) and is necessary to ensure identification[16]. Finally, Assumption 3.A.4 gives some (rather) weak finiteness conditions about the underlying interaction effects functions.

Proposition 3.1. *Let $d_1 = 0, (d_2,d_3) \in \{\bar{\mathbb{R}}_+^2 : d_3 \geq d_2\}$.[17] Under Assumptions 3.A.1-3.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 3.A are identified from the joint distribution of $\{T_1,T_2,T_3\}|\{d,x\}$.*

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

[16] 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.

[17] We define $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$. The statement $d_j = \infty$ implies that the corresponding subject never enters the risk process. Note that, for a maximal group size of M, all groups in the sample of size $J < M$ can be expressed by setting $d_{J+1} = ... = d_M = \infty$.

at different calendar dates, after which they become at risk of using drugs. Here, it should be pointed out that we do not exploit full variation in d_j across members.[18] 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 within groups complicate the identification of Model 3.A, given that the time until the first transition within a group occurs can no longer 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, in assuming that the variation in entry dates is exogenous, the required variation in covariate effects can be reduced to one dimension (see Assumption 3.A.1). In the following subsection, we discuss the special case of a common entry date for all members in a group.

3.2.2 Common entry dates

With some parallel-spell data, all group members enter the risk process at the same calendar date $d_1 = d_2 = d_3 = 0$. For example, if 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 of defaulting. We first replace Assumption 3.A.1 with

Assumption 3.A.5. *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 3.A.5 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 (2003a). Assumption 3.A.5 is a little stronger than Assumption 3.A.1. Making use of a stronger requirement stems from the fact that we cannot exploit variation in entry dates in the case of a common entry point for all group members .

[18] 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 3.A. The identification of the corresponding model is trivial by making use of Proposition 3.1.

Proposition 3.2. *Let $d_1 = 0, d_2 = 0, d_3 = 0$. Under Assumptions 3.A.2, 3.A.3, 3.A.4, 3.A.5, the set of functions $\{\Lambda, \Delta_{j,k}, \Delta_{j,kl} :,\ k \neq j \neq l, k < l\}$ and G in Model 3.A are identified from the joint distribution $\{T_1, T_2, T_3\}|\{x\}$.*

A simple two-spell version of Model 3.A with $d_1 = 0, d_2 = 0$ is formally introduced in Abbring and Heckman (2007), with the authors suggesting an identification strategy.

3.2.3 Relaxing the proportionality assumption

In this subsection, we consider a set of conditions under which the proportionality assumption in Model 3.A may be dropped. For this purpose, we require some of the covariates to vary over time. More precisely, consider the covariate process $\chi_j : \mathbb{R}_+ \to \mathbb{X} \subseteq \mathbb{R}^\mu$, 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 besides 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, which leads to the following multiple-spell duration model

Model 3.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

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

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

Furthermore, we employ the following assumptions

Assumption 3.B.1. *The function $\tilde{\lambda} : \mathbb{R}_+ \times \mathbb{X} \to (0, \infty)$ 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 3.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 3.B.3. *The function G is such that it does not depend on x and d .*

Assumption 3.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*

$$
\begin{aligned}
\Upsilon_{j,k}(t|s,\mathcal{N}_k,\chi_j) &:= \int_{\max\{0,s\}}^{t} \tilde{\lambda}(\omega,\chi_j(\omega))\delta_{j,k}(\omega|s,\mathcal{N}_k,\chi_j(\omega))d\omega,\\
\Delta_{j,k}(t|s,\mathcal{N}_k,\chi_j) &:= \int_{0}^{t} \delta_{j,k}(\omega|s,\mathcal{N}_k,\chi_j(\omega))d\omega,\\
\Upsilon_{j,kl}(t|s,y,\mathcal{N}_{kl},\chi_j) &:= \int_{\pi(s,y)}^{t} \tilde{\lambda}(\omega,\chi_j(\omega))\delta_{j,q}(\omega|\min\{s,y\},\mathcal{N}_q,\chi_j(\omega))\delta_{j,kl}(\omega|s,y,\mathcal{N}_{kl},\chi_j(\omega))d\omega,\\
\Delta_{j,kl}(t|s,y,\mathcal{N}_{kl},\chi_j) &:= \int_{0}^{t} \delta_{j,kl}(\omega|s,y,\mathcal{N}_{kl},\chi_j(\omega))d\omega
\end{aligned}
$$

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 3.B.1 deals with measurability and finiteness conditions of the (integrated) generalized baseline hazard. Assumption 3.B.2 ensures that two different covariate paths exist that agree on an open interval. Note that the latter can be satisfied by simply considering a single covariate that will meet the condition of Assumption 3.B.2. In contrast to Assumption 3.A.3, Assumption 3.B.3 does not impose any conditions on the first moment of the unobserved terms, due to the presence of time-varying covariates (see Heckman and Taber, 1994; Brinch, 2007). Finally, Assumption 3.B.4 is similar to Assumption 3.B.4 and is concerned with finiteness conditions of the underlying functions.

Proposition 3.3. *Let $d_1 = 0, (d_2,d_3) \in \{\bar{\mathbb{R}}_+^2 : d_3 \geq d_2\}$. Under Assumptions 3.B.1-3.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 3.B are identified from the joint distribution of $\{T_1,T_2,T_3\}|\{d,x\}$.*

Note that, in contrast to Model 3.A the finiteness of the first moment of the unobserved terms is not necessary due to the presence of time-varying covariates (see Heckman and Taber, 1994; Brinch, 2007).

As in the case with different entry dates, we can also relax the proportionality assumption in the setting with common entry dates. 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 3.B.5. *It holds* $\tilde{\lambda}(t, \chi_j) = \tilde{\lambda}(t, \zeta_j) \; \forall \; t > 0, \chi_j \in \mathcal{P}_\chi, \; \zeta_j \in \mathcal{P}_\zeta, j = 1, 2, 3.$ [19]

Assumption 3.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 3.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 3.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 is the existence of a certain member-specific characteristic that only directly affects the member but not the other group members.

Proposition 3.4. *Let* $d_1 = 0, d_2 = 0, d_3 = 0$. *Under Assumptions 3.B.1-3.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 3.B are identified from the joint distribution of* $\{T_1, T_2, T_3\}|\{d, x\}$.

3.3 Empirical application

In the following, we present our empirical Application. First, we introduce our data set, before subsequently discussing the estimation method and finally presenting our results.

3.3.1 Data

In our empirical study, we use data from the National Longitudinal Survey of Youth 1979 (see National Longitudinal Surveys Handbook, 2005, for an introduction), which was established in an effort 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

[19] We keep the notation simple here. Specifically, we use the same notation for the extended baseline hazard, $\tilde{\lambda}$, although this function does not depend on the individual characteristics of the other group members as in the case with varying entry dates.

at the time of the first round in 1979, were also included in the survey.[20] Accordingly, 12,686 respondents are included, living in 7,490 unique households. We restrict attention to single-respondent households, as well as those with more than one respondent where the respondents are siblings (blood-related and not blood-related) and grew up together in their parents' home.[21] 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 month they started using marijuana/hashish for the first time in their life. 5,578 respondents report the month and year, with 3,723 never having used up to the interview date in 1984. Based on this and using information on the respondents' birth dates, 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 the surveys of 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, enabling us to 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.

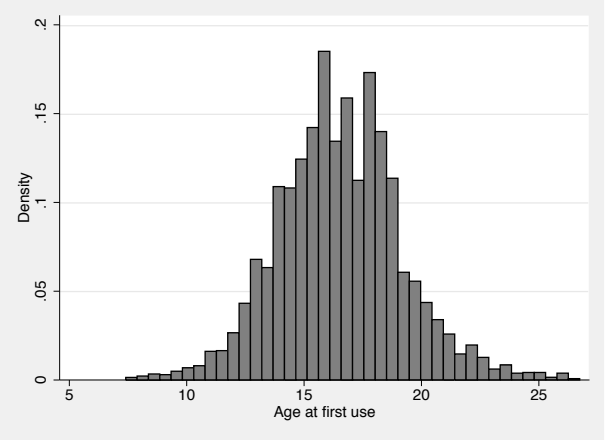
We combine the detailed information on monthly marijuana-use from January 1979 to July 1984 with annual information on the amount and frequency for all relevant survey years. Based on this, we can select the cases in which a first-time use is followed by a long-term change in drug-use behavior.

The resulting distribution of transition times pooled over all household members is

[20] 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.

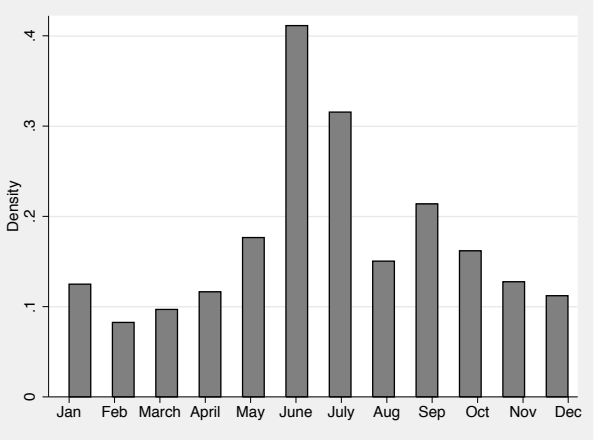
[21] In the majority of all households selected this way, the siblings grew up living with both biological parents. We can observe the time when individuals leave their parents home and the reason for this move (e.g. divorce of the parents). In the analysis of social interaction effects we account for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

Figure 3.2: Distribution of age at first marijuana use



Note: Distribution reflects the full sample pooled over all siblings in the household.

Figure 3.3: Distribution of the month at first marijuana use

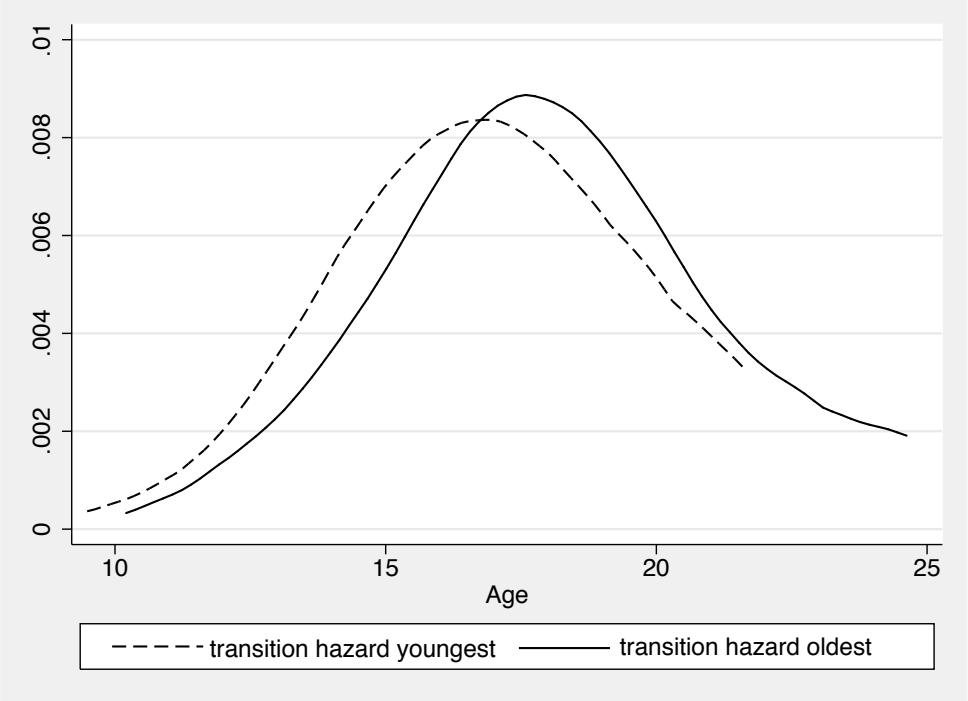


Note: Distribution reflects the full sample pooled over all siblings in the household.

presented in Figures 3.2 and 3.3, with the left figure showing the distribution of age at transition. Before the age of seven, only very few transitions occur. We drop those cases from the sample[22] 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 of June and July, during which American teenagers often go to summer camp and/or spend much time outside. In September, the number of transitions increases again, when 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 3.4 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 the oldest and youngest siblings in the households, with younger siblings transitioning at an earlier age compared to their older siblings. This effect could be driven by observable or unobservable characteristics such as the cohort or character traits, which differ between the oldest and youngest sibling. An alternative explanation is the existence of positive social interaction effects, whereby younger siblings experience the transitions of their older siblings, thus making them more likely to transition at an early age. To determine the

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

Figure 3.4: Transition hazards of the youngest and oldest sibling in the household from two separate Cox regressions



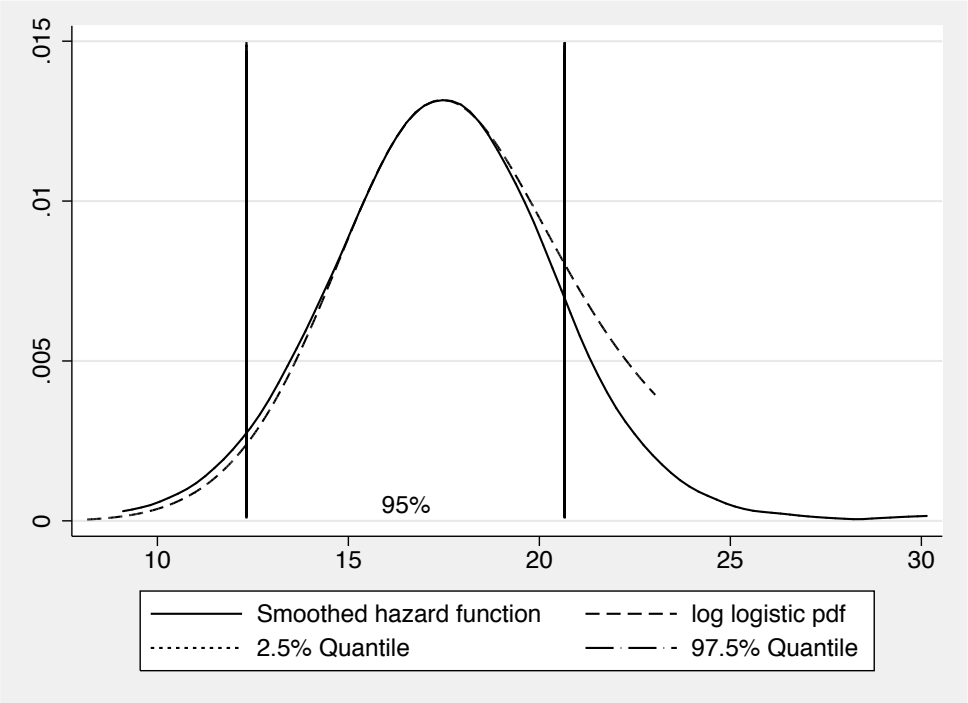
Note: Estimated transition hazard rates of two single spell Cox proportional hazards models using data on first-time marijuana use from the NLSY79; cohorts 1957-1964 of households with at least three siblings.
Dashed curve: Estimated on sample of the youngest sibling in each household; Solid curve: Estimated on sample of the oldest sibling in each household.

source of this difference, we now proceed with the estimation of our multivariate duration model with social interaction effects introduced in the previous section.

3.3.2 Maximum likelihood estimation

Model 3.A provides a general framework of a multiple-spell model with interaction effects, allowing to specify a variety of models fitting different applications. In order to estimate a model using data on first-time marijuana use, we specify functional forms of λ_j, ϕ_j, δ_j and G. Accordingly, the semi-parametric form of Model 3.A is reduced to a model with a finite set of parameters that can be estimated using standard maximum likelihood methods.
Figure 3.5 (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 model. This function has a positive range and is able to approximate the shape of

Figure 3.5: Parametric approximation of baseline hazard function



Note: Estimated baseline hazard functions using data on all households with at least three siblings from the NLSY79 (cohorts 1957-1964) on first-time marijuana use.
Dashed curve: Cox proportional hazard model; Solid curve: 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 and a shared frailty term on the household level.

the baseline hazard estimated by the more flexible Cox model (see Figure 3.5). In the main model specification, we assume proportionality of the covariate effects (Model 3.A), leading 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 $j > 1$, $q = 1, 2$.
Furthermore, we specify the social interaction effect function δ_j with several multiplicative terms, each representing the influence of an experienced transition of a sibling. For a sibling

j living in a household i members at time t, this yields

$$\delta_j(t|T_{i,-j}, x_{ij}(t)) = \prod_{k \in -j} \delta_{j,k}(t|T_{i,jk}, x_{ij}(t))^{I(t>T_{i,jk})},$$

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

and with $T_{i,-j} := \{T_{i,jk} : k \in -j\}$, $-j := \{k \in J_i : k \neq j\}$, $\gamma_k = \gamma_{oldest}$ 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 first term captures unobserved heterogeneity of the hazard rates across households, while the second 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 two types of 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 $P(V_i^{ind} = m_1^{ind}) = p^{ind}$. Accordingly, the distribution of V_{ij} that is the sum of V_i^{sh} and V_{ij}^{ind}, has four mass-points. Note that the term V_i^{sh} that is shared across members of the same household generates a correlation between terms V_{ij} and V_{ik} $\rho = \frac{\sigma_{sh}^2}{\sigma_{sh}^2 + \sigma_{ind}^2}$, where $\sigma_{sh}^2 = Var(V^{sh})$ and $\sigma_{ind}^2 = Var(V^{ind})$.

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

$$\theta_j(t|\{T_{i,-j}\}, x_{ij}(t), V_{ij}) = \frac{\frac{\alpha_{2,j}}{\alpha_{1,j}}(\frac{t}{\alpha_{1,j}})^{\alpha_{2,j}-1}}{(1 + (\frac{t}{\alpha_{1,j}})^{\alpha_{2,j}})^2}\ e^{\beta_{0,j} + \beta' x_{ij}(t)}$$
$$\cdot \prod_{k \in -j} \delta_{j,k}(t|T_{i,jk}, x_{ij}(t))^{I(t>T_{i,jk})}\ \ e^{\gamma_{count} \sum_{l \in -j} I(t>T_{i,jl})}\ \ V_{ij} \quad (3.4)$$

$$S_j(t|\{T_{i,-j}\}, x_{ij}(t), V_{ij}) = exp(-\sum_{l \in -j} I(T_{i,jk} > 0)[\tilde{\Lambda}_j(T_{i,jl}, x_{ij}(t)) - \tilde{\Lambda}_j(\max_{k \in -j_l}\{0, T_{i,jk}\}, x_{ij}(t))]$$
$$\cdot\, \delta_j(T_{i,jl}|\{T_{i,-j}\}, x_{ij}(t))\ V_{ij} \quad) \qquad (3.5)$$

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 in each household are only observed up to a common calendar date at which the interview is conducted in 1984. We denote the resulting vector of censoring points as $c_i = (c_{i1} \ ... \ c_{iJ_i})$.[23] With this information, we can construct the likelihood contribution of a household i.

$$
\begin{aligned}
&Ł(t_i, c_i, x_i; \alpha, \beta, \gamma, m, p) \\
&= \int_0^\infty (\prod_{j \in J} \int_0^\infty \theta_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), V_{ij})^{I(c_j=0)} S_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), V_{ij}) dG^{ind} \ ) dG^{sh} \\
&= \sum_{q=1}^{2} \sum_{q_1=1}^{2} ... \sum_{q_{J_i}=1}^{2} \prod_{j \in J_i} [\ \theta_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), m_q^{sh} + m_{q_j}^{ind})^{I(c_j=0)} \\
&\qquad \cdot S_j(t_{ij}|\{T_{i,-j}\}, x_{ij}(t), m_q^{sh} + m_{q_j}^{ind})]. \qquad (3.6)
\end{aligned}
$$

3.3.3 Results

We estimate our model of first time use of marijuana based on the likelihood specification described in Section 3.3.2. 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 3.1. Model I represents a simple framework with covariates and a basic specification of social interaction effects, yet without accounting for unobserved characteristics (no correlated effects: $\sigma_{sh}^2 = \sigma_{ind}^2 = 0$). 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 his/her oldest sibling starts to use marijuana. $\gamma_{younger}$ measures the effect if one of the younger siblings starts with this habit. 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,

[23] 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).

we still find a highly significant positive effect of a transition of the oldest sibling in the household, however we do not find a significant effect for the transition of a younger sibling. In Model III, we allow for additional flexibility of the social interaction effect functions and find that females are more strongly influenced 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 the social interaction effects reflect the estimated effects of a dummy that has a value of one if the sibling who starts using drugs and the sibling who is affected by this transition are both of the same gender. We do not find evidence for this effect of having the same gender.

The estimated probabilities and mass points described in Section 3.3.2 imply variances σ^2_{sh}, σ^2_{ind} of the two distributions G^{sh}, G^{ind} and correlation ρ between the unobserved heterogeneity terms of two group members V_{ij} and V_{ik}. The parameters are reported in the under 'Correlated Effects' in Table 3.1. We find evidence for unobserved heterogeneity across households ($\sigma^2_{sh} \approx 0.1$), yet not across siblings within households ($\sigma^2_{ind} \approx 0.01$) 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 that the oldest sibling in a household influences his younger siblings in terms of his marijuana use. However, we do not find evidence for an effect of a younger sibling's transition. 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 also explain a substantial part of the dependence in marijuana use behavior.

Table 3.1: Estimation results of Model 3.A using sibling data on first time marijuana use

	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	-	-	.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	

Note: Estimation of three specifications of Model 3.A described in Section 3.3.2 using data from the NLSY79 on first time marijuana use of siblings in American households with at least three siblings growing up together. Model I: Basic specification of social interaction effects; Model II: Basic specification of social interaction effects with discrete distribution of unobserved household characteristics; Model III: Flexible specification of social interaction effects with discrete distribution of unobserved household characteristics. Estimates are reported for $\beta_{0,1}, \beta, \gamma_k, \gamma_x, \gamma_{int}, \sigma^2_{sh}, \sigma^2_{ind}$ and ρ, whereby *, **, or *** reflect a 0.1, 0.05, or 0.01 significance level.

3.4 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 that there can be strong asymmetries across different household 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 social multiplier effects over time. This allows predicting the impact of public policies, depending on which members are initially targeted.[24]

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 limitations of yearly survey data, which is primarily used in studies of social interactions. However, register data and data of members of online platforms constitute an increasingly important data source, providing very detailed information on the timing of actions, making methods exploiting this information increasingly valuable to empirical research.

[24] At this point, in our empirical analysis we restrict attention to the identification of key members in a group and leave the calculation of social multiplier effects over time to future work.

3.A Appendix

3.A.1 Notation

Before proceeding, we introduce some notation and conventions used throughout this Appendix. The symbol G with some (double) subscript refers 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 $j = 1, 2, 3$,

$$Q_{T_j}(t|d, x) := \mathbb{P}(T_j > t, T_j + d_j < \min_{k \in \{1,2,3\} \neq j} (T_k + d_k)|d, x). \tag{3.7}$$

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 $j \neq k$,

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

3.A.2 Proof of Proposition 3.1

The proof of Proposition 3.1 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

$$\mathbb{P}\left[T_1 > t | d, x\right] = \mathcal{L}_{G_1}(\phi_1(x)\Lambda_1(t)). \tag{3.8}$$

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

$$\frac{\partial}{\partial t} Q_{T_2}(t|d, x) = \phi_2(x)\lambda_2(t)\mathcal{L}^{(2)}_{G_{12}}(\phi_1(x)\Lambda_1(t + d_2), \phi_2(x)\Lambda_2(t)), \tag{3.9}$$

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

$$\lim_{t\to 0, d_2 \to 0} \left[\frac{\partial}{\partial t} Q_{T_2}(t|d, x) \Big/ \frac{\partial}{\partial t} Q_{T_2}(t|d, x^*)\right] = \phi_2(x), \tag{3.10}$$

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

$$\mathbb{P}\left[\bigcap_{j=1}^{2} (T_j + d_j > t + d_2) \Big| d, x\right] = \mathcal{L}_{G_{12}}(\phi_1(x)\Lambda_1(t + d_2), \phi_2(x)\Lambda_2(t)). \tag{3.11}$$

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 (3.11), 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 $x \in \mathbb{X}$,

$$\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G_{12}}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t|t_1 - d_2, 1, x)). \tag{3.12}$$

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 $x \in \mathbb{X}$,

$$\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G_{12}}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)(\Lambda_2(t_1 - d_2) + \Upsilon_{2,1}(t|t_1 - d_2, 2, x))). \tag{3.13}$$

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

$$\begin{aligned}\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G}(&\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t + d_3 - d_2|t_1 - d_2, 1, x), \\ &\phi_3(x)\Upsilon_{3,1}(t|t_1 - d_3, 1, x)).\end{aligned} \tag{3.14}$$

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

$$\begin{aligned}\frac{\partial Q_{T_1}(t_1, t|d, x)}{\partial t_1} = \phi_1(x)\lambda_1(t_1)\mathcal{L}^{(1)}_{G}(&\phi_1(x)\Lambda_1(t_1), \\ &\phi_2(x)\left(\Lambda_2(t_1 - d_2) + \Upsilon_{2,1}(t + d_3 - d_2|t_1 - d_2, 2, x)\right), \\ &\phi_3(x)\Upsilon_{3,1}(t|t_1 - d_3, 2, x)).\end{aligned} \tag{3.15}$$

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 $t > 0, d \in \mathbb{D}$, $x \in \mathbb{X}$,

$$\lambda_j(t+\eta_j) = \frac{\partial Q_{T_j}(t|d,x)}{\partial t}\Big[\mathcal{L}_G^{(j)}(\phi_1(x)\Lambda_1(t+d_3), \phi_2(x)\Lambda_2(t+d_3-d_2), \phi_3(x)\Lambda_3(t))\phi_j(x)\Big]^{-1}, \tag{3.16}$$

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

$$\begin{aligned}\lambda_j(t+\eta_j)\delta_{j,1}(t+\eta_j|t_1-d_j,3,x) = \Big[&\mathcal{L}_G^{(1j)}(\phi_1(x)\Lambda_1(t_1),\\ &\phi_2(x)(\Lambda_2(t_1-d_2)+\Upsilon_{2,1}(t+d_3-d_2|t_1-d_2,3,x)),\\ &\phi_3(x)(\Lambda_3(t_1-d_3)+\Upsilon_{3,1}(t|t_1-d_3,3,x)))\\ &\times\phi_1(x)\lambda_1(t_1)\phi_2(x)\Big]^{-1}\frac{\partial^2 Q_{T_1,T_j}(t_1,t|d,x)}{\partial t_1 \partial t}.\end{aligned} \tag{3.17}$$

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 (3.16), (3.17), by using the definitions of the previous paragraph, imply that we have the following system of two differential equations for almost all $t > 0$

$$\begin{aligned}&\frac{d}{dt}\mathcal{H}(t) = f\left(t, \mathcal{H}(t)\right),\\ &\mathcal{H}(\tau) = \gamma_\tau, \text{ for some specific } \tau \in (0, t_1 - d_3) \qquad \text{(initial conditions)},\end{aligned} \tag{3.18}$$

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

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

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 $x \in \mathbb{X}$,

$$\begin{aligned}\frac{\partial^2 Q_{T_1,T_2,T_3}(t_1, t_2, t|x)}{\partial t_1 \partial t_2} = \mathcal{L}_G^{(12)}(&\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t_2|t_1 - d_2, 1, x),\\ &\phi_3(x)\Upsilon_{3,12}(t|t_1 - d_3, t_2 + d_2 - d_3, 2, x))\\ &\times \lambda_1(t_1)\phi_1(x)\lambda_2(t_2)\phi_2(x)\delta_{2,1}(t_2|t_1 - d_2, 1, x), \qquad (3.19)\end{aligned}$$

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

$x \in \mathbb{X}$,

$$\frac{\partial^2 Q_{T_1,T_2,T_3}(t_1,t_2,t|x)}{\partial t_1 \partial t_2} = \mathcal{L}_G^{(12)}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)\Upsilon_{2,1}(t_2|t_1-d_2,1,x), \\ \phi_3(x)\Upsilon_{3,12}(t|t_1-d_3,t_2+d_2-d_3,3,x)) \\ \times \lambda_1(t_1)\phi_1(x)\lambda_2(t_2)\phi_2(x)\delta_{2,1}(t_2|t_1-d_2,1,x). \quad (3.20)$$

The left hand side of the above equation is observed from the data. By Propositions 3.1 and 3.2, 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 $\delta_{2,1}(t_2|t_1-d_2,3,x)$. □

3.A.3 Proof of Proposition 3.2

The identification strategy we follow is the same as in the proof of Proposition 3.1. 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

$$\lambda_j(t) = \left[\mathcal{L}_G^{(j)}(\phi_1(x)\Lambda_1(t), \phi_2(x)\Lambda_2(t), \phi_3(x)\Lambda_3(t))\phi_j(x)\right]^{-1} \frac{\partial Q_{T_j}(t|x)}{\partial t}. \tag{3.21}$$

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

$$\lambda_j(t)\delta_{j,1}(t|t_1, x) = \Big[\mathcal{L}_G^{(1j)}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)\left(\Lambda_2(t_1) + \Upsilon_{2,1}(t|t_1, x)\right), \\ \phi_3(x)(\Lambda_1(t_1) + \Upsilon_{3,1}(t|t_1, x)))\phi_j(x)\lambda_1(t_1)\phi_1(x)]^{-1} \frac{\partial^2 Q_{T_1,T_j}(t_1, t|x)}{\partial t_1 \partial t}. \tag{3.22}$$

The equations (3.21) and (3.22) imply that we have a system of two differential equations. Following similar arguments to the proof of Proposition 3.1 and employing the result of Lemma 1 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

$$\frac{\partial^2 Q_{T_1,T_2,T_3}(t_1, t_2, t|x)}{\partial t_1 \partial t_2} = \mathcal{L}_G^{(12)}(\phi_1(x)\Lambda_1(t_1), \phi_2(x)\left(\Lambda_2(t_1) + \Upsilon_{2,1}(t_2|t_1, x)\right), \phi_3(x)\Lambda_3(t_3) \\ \phi_3(x)\left(\Lambda_3(t_1) + \Upsilon_{3,1}(t_2|t_1, x) + \Upsilon_{3,12}(t|t_1, t_2, x)\right)) \\ \times \lambda_1(t)\phi_1(x)\phi_2(x)\lambda_2(t_2)\delta_{2,1}(t_2|t_1, x). \tag{3.23}$$

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. □

3.A.4 Proof of Proposition 3.3

Proof of Proposition 3.3. It is straightforward, by Assumption 3.B.2, to show that for all $t \in (t_a, t_b), \chi_1 \in \mathcal{P}_\chi$, and $d \in \mathbb{D}_{2\infty}$,

$$\mathbb{P}\left[T_1 > t | d, \chi_1\right] = \mathcal{L}_{G_1}(\mathcal{L}_{G_1}^{-1}(\tilde{\Lambda}(t, \chi_1)) + C) \tag{3.24}$$

for some $C \neq 0$. Applying the result of Brinch (2007), identification of $\tilde{\Lambda}$ follows. Next, for any $t > 0$, $\chi_1, \chi_2, \chi_3 \in \mathcal{P}_\chi$, and $d \in \mathbb{D}$,

$$\mathbf{P}\left[\bigcap_{j=1}^{3} (T_j + d_j > t + d_3) \Big| d, \chi_1, \chi_2, \chi_3\right] = \mathcal{L}_G(\tilde{\Lambda}(t + d_3, \chi_1), \tilde{\Lambda}(t + d_3 - d_2, \chi_2), \tilde{\Lambda}(t, \chi_3)) \tag{3.25}$$

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 3.1 and thus the details are omitted. □

3.A.5 Proof of Proposition 3.4

Proof of Proposition 3.4. 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 $t > 0, \zeta \in \mathcal{P}_\zeta$,

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

with $\tilde{G}$ being the distribution of the random sum $V_1 + V_2 + V_3$. Applying the result of Brinch (2007), we achieve identification of $\tilde{\Lambda}$ and $\tilde{G}$. Next, we have for $t > 0$ and $\zeta_1, \zeta_2, \zeta_3 \in \mathcal{P}_\zeta$,

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

By Assumption 3.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 3.2 and therefore the details are omitted. □

Chapter 4

Inefficiencies from Strategic Behaviour in International Cooperation: Evidence from the Southern Common Market[1]

4.1 Introduction

Under what circumstances do actors cooperate in international relations and which incorporation mechanism ensures efficient contracting? International institutions often have low capabilities to sanction its members for non compliance with signed agreements. Consequently, international actors have an incentive to free-ride on the cooperation efforts of their partners without contributing themselves by ratifying signed agreements at the domestic level (Keohane, 1984; Oye, 1985; Snidal, 1985). This paper argues that institutions can solve this important cooperation dilemma in ways that have previously not been considered within existing literature on international relations. If treaties simultaneously enter in force in the whole region only once all signatory states have ratified, incentives for free riding no longer exist. The short term incentives for individual defection–a menace to international cooperation, which occupies a prominent place in the literature–cannot arise.

However, we further argue that solving free riding by such institutional means does not

[1]This chapter is joint work with Christian Arnold and Gerard van den Berg

fully eliminate cooperation problems; rather, it merely alters their nature. Participants cooperating under such a regulatory regime may feel tempted to make promises that they do not intend to keep. If costs from non-compliance at the domestic level are sufficiently low, those who contract may wish to reap the benefits of positive public exposure from contracting now, even in the light of reputation costs from not keeping your promise in the future. Under such circumstances, actors may sign contracts that would even involve detrimental effects for them if implemented, because they can easily prevent their effect in the whole region by refraining from ratification. The result is inflationary contracting without any effective policy change in the region.

In this paper, we consider the theoretical and empirical implications of a change in the standard incorporation rules towards a mechanism in which treaties enter in force only once all signatory states have ratified. We present the case of Mercosur as an empirical example of this unusual incorporation mechanism and study the empirical implications using the complete record of ratification durations of all 1,024 regulations adopted in Mercosur between 1994 and 2008. This paper builds on the work by Arnold (2013), who outlines the incorporation mechanism and ratification problems in Mercosur in detail.

A striking feature of Mercosur's ratification record is that only half of all regulations signed at the negotiation table before 2004 have entered into force within five years of their introduction. For the other half, at least one of the member countries has not ratified the regulation after 5 years. Using multivariate duration methods, we find that if actors are exposed to high public or political pressure at the time of signing an international agreement, this significantly prolongs the subsequent ratification process at the domestic level. For example, if there is high public exposure of a regulation in a member country at the time of contracting, political actors may be inclined to signal a cooperative attitude by signing the international agreement, knowing that they can easily prevent the policy change from taking effect in the whole region by prolonging domestic ratification for as long as necessary. As a result, too many regulations are introduced in Mercosur and some members pay the costs of domestic ratification while one member's inactiveness prevents the policy change from taking effect. We argue in this paper that although the unusual mechanism in Mercosur prevents the common problem of free-riding,it in fact causes a new problem of inefficient contracting.

This paper illustrates the effects from the distinct ratification rules according to the

following structure. Following a brief literature review on international cooperation in Section 4.2, we set out with theoretical considerations in Section 4.3. After spotlighting the consequences of ratification rules for the production of collective goods, we present our formal model reflecting the incorporation mechanism in Mercosur.

A second part of the paper subsequently turns to the empirical example. In Section 4.4, we present our data set on regional cooperation in the Southern Common Market, a regional integration scheme between Argentina, Brazil, Paraguay and Uruguay that uses the intriguing incorporation rules presented previously. We provide empirical evidence for our theoretical claims in Section 4.5 with an empirical analysis of the ratification record on all 1,024 regulations adopted between 1994 and 2008, before Section 4.6 concludes.

4.2 Literature

In principle, international cooperation consists of two steps. Governments contract on common future conduct. However, authorities still need to keep their promise from the negotiation tables by altering their behavior in line with the terms of an agreement. The ratification of international agreements reflects a key step towards such compliance. Incorporating international rules into domestic law allows national courts to hold their governments accountable to international commitments (Fearon, 1998; Hathaway, 2007; Keohane, 1984; Koremenos et al., 2001; Scharpf, 1997).

Institutions facilitate cooperation against the backdrop of an anarchic international society. They may shape normative beliefs of actors and alter their behavior (Risse et al., 1999; Checkel, 2005; Finnemore and Sikkink, 1998). Moreover, institutions can change benefits from certain courses of action, thus taking effect on the choice of optimal strategies and mitigating the danger of market failure.

Both stages of international cooperation challenge actors in different ways and institutions can be supportive on both occasions. Self-interested actors only consent to international agreements if they expect a positive reward from doing so. First and foremost, they seek to achieve this by changing the status quo of current conditions of international cooperation. Actors need to solve distributional conflicts and coordinate on one among many possible pareto optimal outcomes (Fearon, 1998; Krasner, 1991; Stein, 1983). Institutions reduce transaction costs during bargaining (Koremenos et al., 2001; Williamson, 1975) by simplifying the exchange of information between actors, allowing contracting parties

to identify efficient bargaining solutions in practice (Young, 1991), and serve as focal points for the choice between multiple pareto efficient bargaining solutions (Garrett and Weingast, 1993). Moreover, institutions offer a stable context for the negotiation of multiple issues at one time. Logrolling and linkage politics may offer pareto efficient bargaining outcomes where separate bargaining would not lead to conclusive results (Stein, 1980; Tollison and Willett, 1979).

Those who sign international agreements may yet be interested in gains from agreements other than a mere change in the status quo. For instance, governments can use international rules as a signal to seeksupport from domestic or transnational advocacy groups (Büthe and Milner, 2008; Hathaway, 2007; Simmons and Danner, 2010; Whitehead and Barahona de Brito, 2005). Moreover, country leaders may even want to create distinct reputations in relation to different audiences. They can try to maintain a country's image as a reliable economic partner at the international level. In the meantime, democratically liable governments need to safeguard national interests and may try to undermine binding international contracts with means that are less visible to foreign partners (Kono, 2006). Once all parties come to an agreement, cooperation problems fundamentally change. Conventionally, international regulations only produce the intended effect if all contracting parties make an effort to adapt the necessary behavior. However, given that such adaption is costly, each contracting party has incentives to save on these investments and hope for others providing the collective good (Keohane, 1984; Oye, 1985; Snidal, 1985). Actors can overcome incentives to renege on cooperative behavior when institutions impose costs on those who defect.

Explicit retaliation compels non-compliers to stick to the cooperative behavior initially promised. International enforcement institutions such as courts or tribunals can increase costs for non-compliance to a remarkable extent (Alter, 2002, 2006; Carrubba, 2005; Downs et al., 1996; Gilligan et al., 2010). Furthermore, domestic institutions can also play an important role in sanctioning their own governments (Hathaway, 2007); for instance, national courts (Hathaway, 2003; Powell and Staton, 2009) or interest groups (Dai, 2006, 2007; Martin, 2008) can hold their own governments accountable. When international organizations harness domestic counterparts, international and domestic institutions may commonly exert such pressure on governments (Koremenos et al., 2001). Institutions permit not only explicit but also implicit ways of retaliation. The reputation

of a government effectively defines whether partners are willing to cooperate in further instances (Fearon, 1998; Guzman, 2008; Snidal, 1985). Furthermore, effective monitoring increases such reputation costs (Kono, 2007).

Despite various explanations concerning how institutions affect actors, theories predict mixed prospects for successful international cooperation. Some argue that states sign only those accords that bear little efforts for adaption with respect to the status quo (Downs et al., 1996). However, contracting under high costs can serve as a screening device to identify those partners who truly wish to implement and comply with a negotiation result (Long et al., 2007; Martin, 2000, 2005; Simmons and Danner, 2010; Von Stein, 2005). The constraining capacity of international institutions favors few, but well complied agreements. The more embracing an international institution's capacity to sanction, the harder it is for the parties to reach a compromise and engage in cooperation (Fearon, 1998; Goodliffe and Hawkins, 2006). Nonetheless, only strong enforcement mechanisms can ascertain effective compliance (Goldstein et al., 2007; Guzman, 2008; Hathaway, 2005).

4.3 How different ratification rules affect strategic considerations for international cooperation

Our model reflects classical game theoretic approaches used for the representation of international cooperation (Snidal, 1985; Stein, 1982). All current models implicitly assume that each of the contracting parties decides about the effectivity of an international agreement on its own. Changing this rule has important consequences for the production of collective goods, as well as an important effect on the strategic structure of international cooperation.

4.3.1 Free riding and threshold provision

First, all parties agree on a contract and commit to comply with its terms. In a second step, actors decide whether or not to keep their promise from the negotiation table. If a government implements the terms of an agreement at the domestic level, it has to bear the costs of adaption. By contrast, if authorities refrain from doing so and prefer to save the costs, it may still be the case that other governments invest sufficient efforts to produce the effects from joint action. However, given that not all contracting partners put in practice

what has been previously agreed, the overall welfare of collaboration declines. Overall, the investment in a change of the status quo only pays off if all participate, although each single partner has incentives to free ride on the efforts of their partners (Olson, 1965; Ostrom, 1990).

Such rules for ratification are not the only ones in international politics. Rather than enacting an international agreement on its own, states may condition the effectiveness on the commitment of partners. In some international regimes, agreements only become effective once all signatory countries have ratified.[2] Under these rules, only the joint efforts of all trigger actual policy change. Accordingly, as long as one of the contracting partners lacks ratification, the benefit from the policy change is not produced; therefore, no motive to free ride on others' efforts arises. In contrast to the rules mentioned above, each individual party has an incentive to invest in policy adaption, because the production of the beneficial collective good only occurs if all participate.[3]

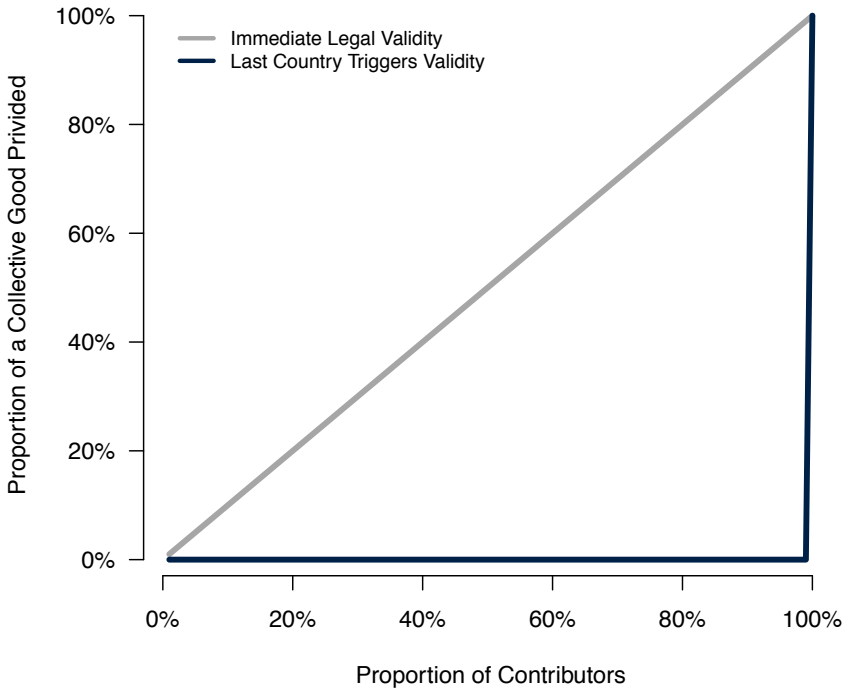
The fundamentally different effects from both rules can be best understood in a public goods framework. The production functions for the collective good differ between the first and second set of rules in important ways, as visualized in Figure 4.1. The horizontal axis depicts the proportion of contributors to a collective good, while the vertical axis charts the ratio of the collective good that is produced for each of the production functions. The grey line represents a production function that leads to the free riding commonly expected in international relations. The more actors that participate in the production of the collective good, the higher the ratio of the good provided. Incentives for free riding exist because parts of the collective good are delivered even when only some cooperate. Defectors are reluctant to invest adaption costs as long as their marginal return from the collective good remains lower.

The black line exemplifies the production function for the latter ratification rule. Collective goods produced with such technologies are know as step-level or provision-point goods within public choice literature (Bagnoli and Lipman, 1989; Kragt et al., 1983; Olson, 1965; Schram et al., 2008). In contrast to the previous case, the collective good is only produced once all comply with their commitments. Consequently, those who do not comply with

[2]Other thresholds for effectiveness of an international agreement are likewise possible.

[3]Under the former rules, enforcement of contracts had to be accomplished at the international level via the threat of implicit and explicit sanctions. Now, domestic courts guarantee governments' effective compliance with agreements beneficial to them.

Figure 4.1: Two production functions for the provision of collective goods



Note: Figure adapted from Ostrom (2003).

their obligations are offered no reward for their reluctance and will be eager to keep their promises from the negotiation table. Of course, this is only true provided that the policy is beneficial to all parties. In the next section, we discuss the consequences of parties secretly opposing the policy change.

4.3.2 Changing the nature of the game

The different rules for ratification fundamentally alter how partners collectively produce a good and thus change the strategic character of interaction (Heckathorn, 1996). With the new rules for ratification, actors no longer face the problem of free riding but may now struggle with inefficient contracting.

To illustrate these consequences, we conceive of a simple game theoretic model whereby two players[4] A and B interact at two stages. In the first round, they may coordinate on a common policy that they promise to later ratify. They make their decision based on the awareness of reputation gains from signaling cooperation today and the anticipated

[4]In our empirical example of Mercosur four players have to decide on international agreements. The consequences of this extension is briefly discussed at the beginning of Section 4.5.

derived utility from their future ratification decision. In the second round, both actors may choose to keep their promise from the negotiation table or rather defect from their commitment. Here, they take their decision in the light of reputation costs that arise from defection and the utility from putting the agreement into practice.

The relationship between these costs and benefits during the two rounds determines the overall dynamic of the game, leading to three scenarios. First, partners consider contracting, but cannot find a bargaining solution in the first round that all partners would agree on. Second, benefits from the new policy are so high that contracting and subsequent ratification is beneficial to all. Finally, the cooperation trap of the game played under these rules for ratification is inefficient contracting. This outcome occurs if the proposed policy change is not beneficial or even harmful to at least one player and reputation gains of contracting are relatively high in comparison to negligible reputation costs from non ratification at the second stage. Under this cost constellation, actors may engage in contracting and benefit from the reputation gains of their signature, knowing that they may easily forestall detrimental policies at low costs. Consequently, international actors sign the international agreement without ratifying it at a later point in time.

We now develop the model in more formal terms. The strategic setting that political actors face can be described by a two-step cooperation game with incomplete information (see Figure 4.2). At the first stage (Contracting stage), the two countries A and B choose to

Figure 4.2: Two player two-step game of international cooperation

	D	C
D	$0, 0$	$-R_A, R_B$
C	$R_A, -R_B$	$X_A + R_A, X_B + R_B$

Stage 1: Contracting

	D	R
D	$-r_A, -r_B$	$-r_A, 0$
R	$0, -r_B$	U_A, U_B

Stage 2: Ratification

Note: Stage 2 of the game is only reached if players cooperate on Stage 1. In particular, an agreement only enters into force, if both countries first sign the international contract (C,C on Stage 1) and then ratify it at the Ratification Stage (R,R on Stage 2).
R_j: reputation gains from contracting of player $j = A, B$; $-r_j$: reputation costs of non ratification; X_j: expected utility from ratification stage; U_j: Utility from policy change being ratified by both players.

introduce a new policy. If neither of the partners cooperates (D,D), both players remain with the status quo and receive no pay-off. In the case where both agree on the introduction of the policy (C,C), each actor $j = A, B$ receives their expected utility from the ratification stage X_j and reputation gain R_j from their signature. The game subsequently proceeds to the second stage (Ratification stage). However, if, for example, country B defects while

A decides to cooperate (C,D), then country A improves its reputation as a cooperative partner (gains R_A), whereas country B has to accept a less benevolent standing (loses R_B).

Once the regulation reaches the Ratification stage, both actors decide whether or not to adhere to the terms of the agreement and enact it into domestic law. While in the first round actors faced the reputation loss $-R_j$ whenever they defect, they now face less severe reputation costs $-r_j$ for not ratifying the agreement.[5] Whenever both players choose to ratify the agreement (R,R), their efforts collectively produce the international policy change with corresponding utility U_j.

International actors may evaluate the utility of the common policy differently. Those who have to ratify an international regulation either cherish the policy ($U_j = H$) or rather disapprove a change in the status quo ($U_j = L$). At the outset, nature draws the utilities U_A and U_B independently from a distribution with $P(U_j = H) = p$ and $P(U_j = L) = 1-p$, $p \in [0, 1]$. Players know the distribution from which the types U_j are drawn, while the realization U_j is private information to player j.

Concluding the description of the game, we impose the following assumptions on the payoffs:

$$R_j, r_j, \rho, H > 0 \tag{3.1}$$

$$-r_j > L \tag{3.2}$$

with ρ being the discount factor for the payoff at the ratification stage. Assumption (3.2) formalizes the statement that a country of type L will be more strongly harmed by the policy change than from the reputation loss $-r_j$ that it would suffer from defecting at the second stage. As a result, a country of type L always chooses not to ratify the regulation as long as the policy would enter into force otherwise.

To understand the dynamic of the game, we first consider possible equilibria at the ratification stage, given that contracting by both players has taken place. In a second step, we subsequently turn to the overall dynamic of the game and consider actors anticipating

[5] As such, defection at the second stage corresponds to choosing an incorporation time of ∞. Since the strategy of infinitely delaying incorporation can only be indirectly observed by the other country, it is difficult to publicly condemn this behavior. Instead, political actors have to face a less severe reputation loss within the political system and against lobbies supporting the law.

the outcomes at the ratification stage during contracting.

Under assumptions (3.1) and (3.2), different combinations of types H and L lead to three different action situations at the ratification stage, all of which are characterized by different Nash equilibria. If a player benefits from a the policy change $U_j = H$, irrespective of the partner's strategy, a player's best response is always quick ratification. Thus, if both players benefit from the policy change $U_A = H, U_B = H$, it will be ratified as quickly as possible by both players (R,R). If, on the other hand, player A is of type L while player B is of type H and thus ratifies quickly, the best response of L is to defect (assumption 3.2), while in response B still has an incentive to ratify quickly to avoid reputation loss $-r_B$ (D,R). The reversed scenario occurs for $U_A = H, U_B = L$ respectively (R,D). As a result, opposed preferences lead to equilibria whereby one contracting partner does not fulfill their obligation and hence the policy change never enters into force (inefficient contracting). Finally, if both countries are of type $U_A = L, U_B = L$, two equilibria may arise. If A chooses to defect, the best response of B is quick ratification to avoid reputation costs $-r_B$.[6] In turn, the best response of A is to defect (D,R). The second equilibrium occurs in case of the reversed combination of strategies (R,D). In the following, we assume that the two players will coordinate on either one of the two equilibria with probability $\frac{1}{2}$. Similar to the previous case, the policy change never enters into force (inefficient contracting).

Actors who consider international cooperation during the first round anticipate the equilibrium outcomes from the ratification stage. From the perspective of player j, successful contracting leads to the following expected utilities:

$$X_j = \begin{cases} \rho p H & \text{if } U_j = H \\ -\rho \frac{1}{2}(1+p) r_j & \text{if } U_j = L. \end{cases}$$

Considering the structure of payoffs in figure 4.2, it is easy to understand under which conditions both actors are eager to introduce a new policy. As long as $X_A + R_A > -R_A$ and $X_B + R_B > -R_B$, both players' best response to contracting is to sign the contract themselves (C,C). For players with $U_j = H$, this condition is always satisfied: $\rho p H + R_j > -R_j$. Thus, if the policy is beneficial to both parties, actors contract at the first stage

[6] In order to avoid unnecessary reputation loss, the best response of B to defection is quick ratification. This can be interpreted as a minor form of free-riding on the defection efforts of the other player in order to avoid reputation loss $-r_B$.

and the regulation is subsequently quickly ratified by both. This theoretical implication is also in line with our empirical findings presented in Section 4.5. Regulations known to be beneficial to all members are on average more quickly ratified by all member countries compared to regulations with unknown preference distributions. Note that in standard public goods games, the incentive to free-ride would cause a less favorable ratification record for a mutually beneficial policy change.

A country with $U_j = L$ will only choose to agree to introducing the new policy if the public reputation loss from defecting now is larger than the expected (internal) reputation loss from defecting at the ratification stage:

$$\begin{aligned} & -\rho\frac{1}{2}(1+p)r_j + R_j > -R_j \\ \Leftrightarrow \quad & \underbrace{\frac{1}{2}\rho(1+p)r_j}_{X_{j,L}} < 2R_j \end{aligned} \tag{3.3}$$

Condition (3.3) reveals the potential inefficiency resulting from the legislative mechanism in Mercosur. If inequality (3.3) holds, a country of type L prefers the risk of an expected future reputation loss of not ratifying $-\rho\frac{1}{2}(1+p)r_j$ to the effective reputation costs $-2R_j$ of not signing the agreement during the negotiations. As long as incentives to contract today are sufficiently high to outweigh the expected losses tomorrow, actors who are secretly opposed to the regulation ($U_j = L$) prefer to publicly sign the international agreement, knowing that they benefit immediately and can meanwhile prevent the policy at little cost tomorrow.

Table 4.1 summarizes the equilibrium outcomes of the two-step cooperation game depending on the utilities derived from the policy change $U_A, U_B \in \{H, L\}$, expected utilities $X_{B,L}, X_{A,L}$ and reputation gains from contracting R_A, R_B. Note that the three highlighted cells in Table 4.1 $\{U_A = L, \frac{1}{2}X_{A,L} < R_A, U_B = H\}$, $\{U_A = H, U_B = L, \frac{1}{2}X_{B,L} < R_B\}$ and $\{U_A = L, \frac{1}{2}X_{A,L} < R_A, U_B = L, \frac{1}{2}X_{B,L} < R_B\}$ result in the introduction of the regulation, although the policy change never enters into force, given that one of the contracting partners fails to keep their promise. We claim in this paper that the slow ratification behavior observed in Mercosur can be partly explained by this process of inefficient contracting.

In the dataset of Mercosur, we only observe the regulations that have passed the contracting stage, i.e. whereby all members have signed the contract. Accordingly, in the

Table 4.1: Summary of equilibrium outcomes

		Typ H	Typ L	
			$\frac{1}{2}X_{B,L} < R_B$	$\frac{1}{2}X_{B,L} \geq R_B$
Type H		(C,C) (R,R)	**(C,C)** **(R,D)**	(C,D) -
Type L	$\frac{1}{2}X_{A,L} < R_A$	**(C,C)** **(D,R)**	**(C,C)** **(R,D)** or **(D,R)**	(C,D) -
	$\frac{1}{2}X_{A,L} \geq R_A$	(D,C) -	(D,C) -	(C,D) or (D,C) -

Note: The equilibrium outcomes depend on the realization of preferences $U_A, U_B \in \{H, L\}$, expected utilities $X_{B,L}, X_{A,L}$ and reputation costs R_A, R_B. The first row of each cell represents the combination of strategies of player A and B at the Contracting stage. The second row reflects the strategies at the Ratification stage.

empirical analysis we are interested in identifying the factors that increase the likelihood of the occurrence of inefficient contracting conditional on the event that the regulation has been signed by all members. In order to derive an expression for this conditional probability, in the following we consider R_A (R_B) in inequality (3.3) as a positive random variable that varies across regulations independently from U_A, U_B and R_B (R_A). Furthermore, we assume that a component E_j of R_j is observable $R_j = E_j + R_j^{unobs}$. Here, E_j denotes a set of observable influences, such as the public and political environment at the time of contract signing, that directly affect the potential reputation gain (or loss) R_j ($-R_j$) of (not) signing the international agreement. We denote by $p_{X_j}(R_j) = P(\frac{1}{2}X_{j,L} < R_j) \in (0, 1]$ with $\frac{\partial p_{X_A}(R_A)}{\partial R_A} > 0$ the probability of a sufficiently high realization of R_j, according to which, if country j is of type L, it prefers to publicly sign the regulation despite secretly opposing it. Furthermore, we assume $p \geq \frac{1}{2}$.[7]

Now, we can derive an expression for the probability of country A not ratifying the regulation given that both A and B have signed it at the contracting stage:

$$\begin{aligned} P(D_A|C_A, C_B) =& \frac{P(D_A, C_A, C_B)}{P(C_A, C_B)} \\ =& \frac{p(1-p)p_{X_A}(R_A) + \frac{1}{2}(1-p)^2 p_{X_A}(R_A)p_{X_B}(R_B)}{p^2 + p(1-p)[p_{X_A}(R_A) + p_{X_B}(R_B)] + (1-p)^2 p_{X_A}(R_A)p_{X_B}(R_B)}. \end{aligned} \quad (3.4)$$

[7]Note that the parameter p has a rather abstract interpretation since it reflects the distribution of preferences over the full set of regulations that are discussed (but not necessarily signed) at the negotiation table. We assume here that it is more likely that a country benefits from a regulation considered at the negotiation table than to be harmed by it.

Taking the derivative with respect to E_A yields:

$$\begin{aligned}
&\frac{\partial P(D_A|C_A, C_B)}{\partial E_A} \\
=&\frac{p(1-p)p_{X_A}(R_A)(1-P(D_A|C_A, C_B)) + (1-p)^2 p_{X_B}(R_B))\frac{\partial p_{X_A}(R_A)}{\partial E_A}(\frac{1}{2} - P(D_A|C_A, C_B)}{P(C_A, C_B)} \\
>& 0. \qquad (3.5)
\end{aligned}$$

The last inequality follows directly from $\frac{\partial p_{X_A}(R_A)}{\partial E_A} > 0$ and $(3.4) < \frac{1}{2}$. Thus, the theoretical model implies that if the public and political environment places increasing pressure on international decision makers at the negotiation table, inefficient contracting is more likely to occur. Indeed, we find empirical evidence of this effect in the ratification behavior of international actors in Mercosur. Note that the conditional probability in (3.4) is closely connected to the quantity that we model in the empirical section, where we specify a statistical model for the ratification durations in Mercosur for all regulations that have been signed by all signatory members.

4.4 The Southern Common Market

We now present the empirical case of the Southern Common Market, a regional integration scheme between Argentina, Brazil, Paraguay and Uruguay that uses the regulatory regime explained above for the ratification of its regulations. Interestingly, its member countries duly ratify only two-thirds of all policies, but are apparently reluctant to fulfill their obligations in the extant cases. First, we provide a brief overview over the institution of Mercosur, before subsequently presenting our ratification data set in detail.

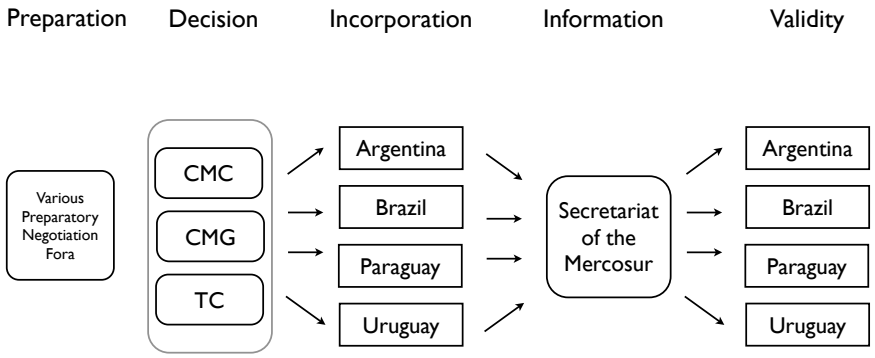
4.4.1 Southern Common Market and the rules for ratification

Figure 4.3 depicts the institutions relevant for regional cooperation and explains the rules for ratification as established in the Protocol of Olivos (POP) from 1994. According to this treaty, the member states unanimously take decisions in one of the three decision bodies: the Common Market Council (CMC), the Common Market Group (CMG) or the Trade Commission (TC) (Bouzas and Soltz, 2001; Lavopa, 2003; Ventura and Perotti, 2004). Each member state incorporates Mercosur regulations into the respective domestic

juridical system and subsequently communicates successful ratification to the Secretariat. However, according to Article 42 POP, a regulation is not yet legally in force; only 30 days after every member state successfully ratified does a regulation develop its binding character in all Mercosur countries. When one single state complies with the legal term of an agreement in the Southern Common Market, this member country prepares the effect of a Mercosur regulation, although it does not yet trigger it. The institutional design ensures that rules enter into force in all countries at the same time.[8]

In view of these intriguing rules for ratification, Mercosur represents an apposite case to observe the dynamics of the theoretical model in practice. Reputation costs from

Figure 4.3: Regulatory framework for International Cooperation in Mercosur



non ratification are particularly low, while for many regulations the public, political or institutional environment during contracting ensures high rewards from contracting alone. This renders political actors in Mercosur particularly susceptible to inefficient contracting. Moreover, we can empirically validate whether the theoretical prediction of no free-riding in the case of mutually beneficial policies is empirically confirmed in Mercosur.

During ratification, the threat of implicit sanctions such as reputation loss is lower in Mercosur in comparison to other international regimes, given that information concerning the status quo of ratification is not publicly available. While the secretariat collects information on successfully incorporated cases, access to this database remains restricted, with neither interest groups nor citizens able to gain any insight. The Secretariat only

[8] Alejandro Pastori, who was the legal adviser of the Uruguayan Foreign Minister during the negotiations of the Protocol of Ouro Preto in 1994, compared this procedure to a swimming pool. All swimmers would step close to edge. Only if everyone was ready, all would jump at the same time. (Interview in Montevideo, April 2009)

provides information for individual regulations, thus making it hard for public actors to hold their governments accountable.[9] Mercosur's procedures are similar regarding information vis-Ã -vis partner governments. While meetings and minutes exist that intend to brief about ratification endeavors in particular policy fields, the overall ratification record remains officially under disclosure.

Typically, effective dispute settlement mechanisms allow for overcoming cooperation problems with explicit sanctions (Downs et al., 1996; Fearon, 1998; Yarbrough and Yarbrough, 1997). Mercosur's institutional provisions remain comparably limited in this respect and largely intergovernmental in character (Lenz, 2012; Malamud, 2005; Pena and Rozemberg, 2005). According to the categorization for dispute settlement in international trade from Yarbrough and Yarbrough (1997), Mercosur's dispute settlement system would file in the second weakest category. While a third party may settle a dispute and advice retaliatory measures, it neither implements nor controls implementation (Bouzas et al., 2008, 100). In the terms of Keohane et al. (2000), Mercosur's dispute resolution displays a moderate level of legalization (Krapohl et al., 2009). Despite the existence of institutions for conflict adjudication, their independence, access and legal embeddedness do not provide for an effective enforcement of non-compliance.

4.4.2 Measuring actors' ratification behavior

Mercosur is not only an intriguing case due to its particular rules for ratification, but also given it empirical ratification record. We use information relating to the success and duration of ratification in the four member countries concerning all 1,024 regulations adopted between 1994 and 2008 (Arnold, 2013).[10] Pooling this information, our data set contains 3,560 data points and offers insight into the ratification success and duration as measured at the end of 2008.

Mercosur's members do not incorporate all regulations into the respective domestic legal bodies, with Table 4.2 offering insight into the dimension of this issue. Half of the regulations introduced in Mercosur between 1994 and 2003 have not entered in force five years after their introduction date due to at least one of the member countries failing

[9] The Secretariat's website allows for querying the status quo of single regulations, only.

[10] While the four countries agreed on 1700 regulations overall, only 1024 of them require active ratification to become legally effective.

Table 4.2: Percentage of ratification success in Mercosur

	Non ratification %
by at least one member	50.4 %
by Argentina	21.7 %
by Brazil	18.8 %
by Paraguay	25.5 %
by Uruguay	26.3 %

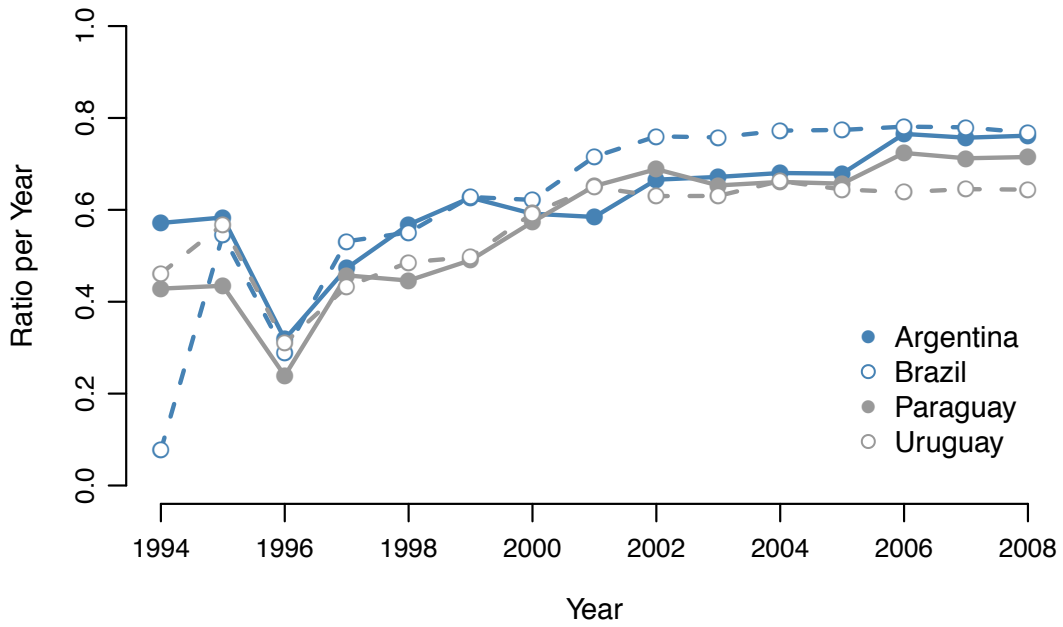
Note: Percentage of regulations introduced between 1994 and 2003 that have not been ratified by at least one of the four Mercosur countries five years after their introduction (below: non ratification percentage by a specific member country).

to ratified the agreement at the domestic level. When examining the potential culprits, we find that the two smaller countries of Paraguay and Uruguay have the worst ratification record in the region with 25% of the regulations not incorporated into national law five years after their introduction.

One natural explanation for the poor overall ratification record in Mercosur is the inefficiency of the national administrative systems in the region. However, within Mercosur, Uruguay is known to have by far the most efficient and least corrupt administration, although it has the worst ratification record in Merosur, closely followed by Paraguay. Thus, there has to be a different explanation for the poor ratification behavior observed. The strategy of not keeping promises from the negotiation table, as outlined in Section 4.3.2, may be one way for the economically and geographically smaller countries of Paraguay and Uruguay to respond to the dominating role of Brazil and possibly also Argentina within the region. The incorporation mechanism of Mercosur provides the smaller countries with the same political power as the larger ones, thus enabling them to prevent policy changes from taking effect in the whole region.

The graphs in Figure 4.4 indicate the development of ratification success over time between 1994 and 2008. The vertical axis shows the success ratio for each country, calculated as all ratified regulations in relation to the overall number of regulations introduced. The introduction of regulations briefly peaked in 1996, causing the success ratio to plummet. The success ratio subsequently increased, reaching a level between 63.0% in Uruguay and 75.7% in Brazil in 2008. Since 1997, while Brazil seems to perform best, Uruguay worst

Figure 4.4: Ratification success in Mercosur over time



Note: Ratification success of regulations in Mercosur for all four member countries. The connected points indicate the incorporation status per country as of the respective year.

and Argentina and Paraguay usually in between. Overall, the ratification success of the four countries follows a similar path over time.

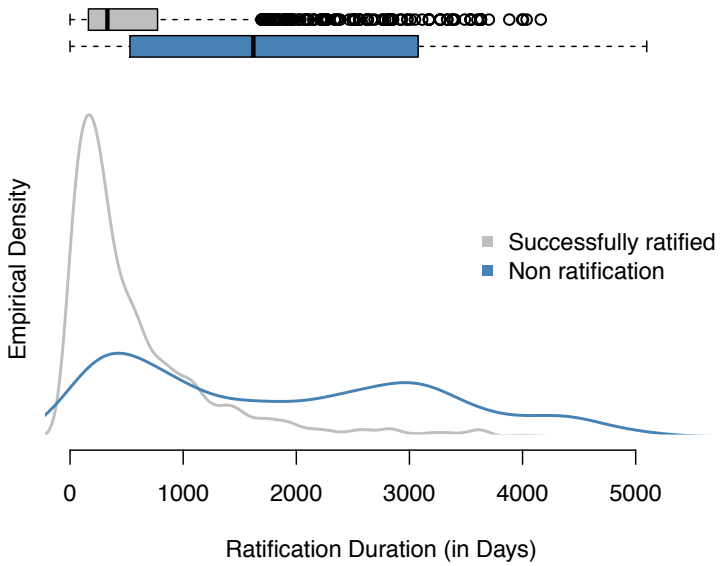
Moreover, in Figure 4.5 we report the empirical density of the successfully incorporated ratification spells (grey curve) and the censored cases (blue curve). The upper graph summarizes the data with the help of boxplots.[11]

We find that successful ratification takes 585 days on average, with the respective 25% quantile at 162 days, the median at 330 days and the 75% quantile at 774 days, with a maximum of 4,161 days. While in 75% of cases, ratification takes less than 26 months (774 days), one-quarter of all policies take between 26 months and 11 years (4,161 days) until they are ratified. Overall, most of the policies find their way into the respective domestic legal system in the four member countries in little time, while the right skewed character of the distribution shows that a substantial number of regulations take considerably more time.

To explain the variation in the success and duration of ratification in Mercosur with multivariate analysis, we measure a number of additional variables, capturing

[11]The black bar within the box stands for the median. The outside edges of the boxes indicate the 25% quantile and the 75% quantile. The whiskers report outliers: any datapoint outside 1.5 times the interquartile range is printed using a black circle.

Figure 4.5: Distribution of ratification durations in Mercosur



Note: Density of the ratification durations for regulations pooled over all countries. The grey line represents ratification durations of successful cases, the blue line stands for ratification durations of all regulations not yet ratified until 2008 (censoring of the observations).

characteristics of the political, institutional and economic context and the regulations themselves. First, we divide Mercosur's policies according to a series of categories (Table 4.3). Among all 3,560 regulations[12], politicians adopted 599 of them in the Common Market Council and bureaucrats signed 2,961 in the Common Market Group and the Trade Commission. In addition, we distinguish between six policy types. 725 regulations relate to the common external tariff, while 800 announce governmental cooperation. In 104 cases, the four members consider Mercosur interna, whereas 87 policies address the internal market, 276 policies define exceptions from the common external tariff and 1,537 concern technical regulations. Moreover, 31 regulations could not be attributed. Finally, not all countries need to ratify all policies; accordingly, Argentina has to incorporate 870 Mercosur rules into domestic law, Brazil 882 policies, Paraguay 929 policies and Uruguay 879 policies.

In addition to these categories, four additional variables measure the political context

[12]Here, the regulations are counted several times depending on the number of signatory member countries.

Table 4.3: Distribution of ratification durations over three categorical variables

	n
Decision Bodies	
CMC	599
CMG and TC	2961
Policies	
Common External Tariff	725
Governmental Cooperation	800
Mercosur Interna	104
Internal Market	87
Tariff Exception	276
Technical Regulations	1537
Others	31
Countries	
Argentina	870
Brazil	882
Paraguay	929
Uruguay	879

(Table 4.4). 'Public support of Mercosur in country' measures citizens' attitudes regarding regional cooperation on a yearly basis.[13] Survey data from the Latinobarometro shows that the population adopts a comparably positive stance towards regional cooperation, on average. Support for Mercosur ranks at 82%, with a standard deviation of 0.06 across all regulations. The variable 'Mercosur presidency' indicates whether the meeting during which the regulation was signed was held in the respective country, which implies a higher media coverage of signed regulations and Mercosur topics in general. Mercosur has a rotating presidency and typically the country chairing the meetings is also the host of

[13]Since there is no question concerning Mercosur which has been asked every year, three different question wordings are used. Despite their difference, all of them address a general consent towards Mercosur. The question which has been asked most frequently is: "Are you in favor or against economic integration in Latin America?". Respondents can answer the question with *very much in favour*, *a little in favour*, *a little against* and *very much against*. We coded the two positive categories as consent and the two negative ones as dissent. In the year 2003, marked with a small star, respondents were asked "Among the institutions that are on the list, please evaluate them in general terms and give them a mark between 0 and 10, where 0 is very bad and 10 would be very good, or else tell me whether you have not heard enough to provide an opinion about: Mercosur". We rescaled the answers to differentiate between those who reject Mercosur and those who do not and coded all responses larger and equal to 5 as positive and those that are smaller than 5 as negative. Finally, in the years 2004, 2006 and 2007, the question was "Treaties on international free trade have a very positive, positive, negative, very negative or no impact at all on your employment opportunities?". Again, we merge the two top categories and the two lower ones.

the negotiations. The variable takes a mean of 0.24 and a standard deviation of 0.43. Furthermore, analyzing media coverage of Mercosur's trade disputes, Gómez-Mera (2009) devises an annual score capturing the perceived level of large salient conflicts between Mercosur's governments, with a mean of 1.69 and a standard deviation of 1.29 ('Conflict level in Mercosur').

Finally, in order to capture the number of domestic actors who have to be coordinated for ratification, we introduce the number of veto players ('Size of opposition in country') of the respective domestic political system. We operationalize this number with the index from Beck et al. (2001). In our sample, this takes a mean of 3.83 and varies with a standard deviation of 1.22.

We further introduce four variables capturing idiosyncrasies of Mercosur's policies. We

Table 4.4: Descriptive statistics of measures of the political context and idiosyncrasies of the regulations

	Mean	s.d.	Min	Max
Political Context				
Public Support of Mercosur in country	0.82	0.06	0.63	0.94
Mercosur presidency	0.24	0.43	0	1
Conflict level in Mercosur (sqrt)	1.69	1.29	0.00	4.30
Size of opposition in country	3.83	1.22	2	6
Policies				
Complexity of Policy	0.29	1.30	-2.41	4.23
Technical annex	0.66	0.47	0	1
Number of references in rreamble (sqrt)	1.83	0.52	0.00	3.87
Overrules Mercosur regulation	0.14	0.35	0	1

approximate the complexity of a regulation with its length and count the number of words and paragraphs.[14] We decompose the variance of regulations' measures for length into the principal component in order to tap the latent complexity dimension. Assigning principle component scores to each regulation, it is possible to interpret complexity on a common scale (Bartholomew et al., 2008; Joliffe, 2002). The mean in our pooled sample is 0.29, with a standard deviation of 1.30. Next, a dummy variable captures whether a policy contains an annex with comprehensive technical details, taking a mean of 0.66 and a standard

[14]We use the software JFreq to count the number of words and rely on handcoding for the number of paragraphs. Both values are log transformed to account for their skewed distributions.

deviation of 0.47. The number of references to xisting Mercosur rules approximates the amount of existing Mercosur legislation that a policy builds upon. Taking the square root to correct for the skew of the variable's distribution, we report an average of 1.83 and a standard deviation of 0.52. Finally, another dummy is used to indicate whether a policy overrules existing legislation, with this variable reporting a mean of 0.14 and a standard deviation of 0.35.

4.5 Analyzing the ratification behavior in Mercosur

We argue that part of the reason behind Mercosur's poor ratification record is that the South American actors follow the incentive structure outlined in the theoretical section. However, we have to bear in mind that the two-step game of cooperation only captures the strategic interaction of international actors in a very simplistic way. When analyzing the data on ratification durations in Mercosur, we have to abstract from this simple mechanism in several ways.

4.5.1 Relation of cooperation game to empirical analysis

First, after a regulation is signed by the member countries, actors have a larger choice set than merely to defect (never ratify) or cooperate (ratify as quickly as possible). In fact, within the boundaries of the domestic administrative system, political actors in Mercosur can choose practically any ratification duration (measured in days after contract signing). Accordingly, different lengths of ratification durations can be interpreted as different degrees of defectiveness. If a policy change is disadvantageous, political actors may decide not to completely defect but rather postpone ratification as long as possible. Furthermore, many of the observed ratification durations are right-censored due to the end of the window of observation in 2008. We account for this demanding structure of the data by using a multivariate continuous duration approach to jointly model the ratification hazards of the four member countries. This method allows accounting for the continuous duration characteristic of the data and the corresponding problem of right-censoring.
Second, the preferences of international actors towards a regulation may not remain constant but rather change over time. This could be due to a change in the political decision makers themselves, i.e. a shift of power within the domestic decision body or a

change of the political leader due to elections. Additionally, the economic environment that the political actors face may also change over time. Thus, regulations that are beneficial to a country at the time of contract signing may not be beneficial a few years later. We observe several political and economic variables, capturing such changes over time on a yearly basis. The use of a hazard rate approach enables us to account for such time-varying observable changes.

Third, not only two but four countries are involved in international contracting in Mercosur. However, one can think of two potentially opposing parties at the negotiation table, namely the countries in favor of the regulation of interest and those potentially opposing it. Since an increase in the number of international actors also increases the likelihood of at least one actor secretly opposing the regulation, one would expect a higher probability of a defective outcome (no policy change). Thus, the problem of inefficient contracting outlined in the two player game is magnified in a game with four international actors.[15]

It is important to understand that the existence of multiple equilibria in the ratification period prevents the full identification of the underlying preferences from the observed ratification behavior. For example, if country A cooperates in the ratification period, while country B defects, this does not automatically reveal that the regulation is beneficial to country A, since it is also possible that A anticipated that B will defect and consequently decided to free-ride on B's defection effort. Thus, with the given set of assumptions, identification of the underlying preferences is not possible from the data.

Furthermore, when studying the ratification behavior of international contracting, one typically only observes the regulations signed by all members, whereas those regulations already discarded at the negotiation table are usually not documented. Indeed, this is also true for data on Mercosur. Consequently, the distribution of preferences for regulations underlying our dataset is highly selective and does not represent the original distribution from which the utilities/preferences are drawn in the theoretical model. Similarly to the conditional probability (3.4), in the next section we conduct inference conditional on the

[15]Note that, in a game with more than two players the problem of multiple equilibira briefly discussed in Section 4.3.2 will become more complex. For example, a regulation may be discussed addressing an exception to the common external tariff for Brazil that is not beneficial or even harmful to the other three members. If the public or political pressure to cooperate is sufficiently high, Argentina, Paraguay and Uruguay will sign the regulation. Several equilibriums are now possible that result in defection. Each of the three secretly opposing countries could be the one to defect (three possible equilibria).

event that a regulation has been signed by all necessary member countries.
In summary, the objective of the following empirical study is not to identify the parameters and preferences of the game presented in the previous section. Rather, the purpose of the theoretical model is to illustrate the potential for inefficient contracting induced by the incorporation mechanism in Mercosur. In the empirical analysis, we focus on identifying how a certain political, public or institutional environment in a country at the time of contract signing can lengthen the subsequent time until ratification at the domestic level. The results suggest that, conditional on a regulation being signed, high political or public pressure at the time when this decision is made significantly lengthens the subsequent ratification duration in the four member countries. We offer the occurrence of inefficient contracting defined in our theoretical model as an explanation for this empirical finding.

4.5.2 Empirical model

We analyze the observed ratification durations in Mercosur using survival analysis, thus making use of both the success and duration for ratification. In survival models, the unit of analysis enters the observation period with a discrete characteristic and alters this state after a certain amount of time. Here, Mercosur's members adopt a regulation at the Mercosur level and subsequently change its status once they ratify it domestically. If it were for the discrete change only, any ordinal model would suffice. However, event history models seek to answer how long it takes until a certain event occurs, such that duration and status are both of interest (Beck and Katz, 1996; Box-Steffensmeier and Jones, 1997; Johnson and Albert, 1999).
We conceive and model ratification as a process whereby a regulation has a certain probability of being ratified each day, given that ratification has not succeeded up to this point. We model such ratification hazards, making use of the popular proportional hazard assumption that states that the effect of covariates is constant over time and enters the hazard rate multiplicatively. This leads to the following model for the ratification hazard of regulation i in country j at ratification duration t:

$$\theta(t|x_{ijt}) = \lambda(t)\ e^{x'_{ijt}\beta} \qquad (\textit{Model I and II})$$

Here, the function $\lambda(t)$ captures the dependence of theratification hazard on the time t

passed since regulation i was signed by country j. Given that we have little prior knowledge about the shape of this function, we choose a piecewise constant specification. With each additional year since contract signing, a new parameter captures the possibly different level of the ratification hazard.

The effect of covariates is reflected in the vector β

$$x'_{ijt}\beta = c + c_j + y_t + r'_i\beta^r + m'_{ij}\beta^m + \Delta'_{ijt}\beta^\Delta.$$

Here, the covariate vector x_{ijt} includes a constant, dummys for the member countries c_j, dummys for each calendar year (1994,...,2008) y_t and characteristics of the regulation r_i such as measures of complexity of the document signed or the policy field that it addresses. m_{ij} holds the main variables of interest in our analysis specific to the meeting at which the contract is signed. These covariates vary across regulations i and/or over the four countries j and capture the public, political, institutional or economic environment faced by the political actors at the time of contract signing. The variables of interest include a dummy indicating whether the regulation is signed by politicians vs. bureaucrats, the measure of public support of Mercosur in country j at the time, an indicator whether country j holds the presidency of Mercosur and consequently is the host of the meeting, a measure of the conflict level in Mercosur and the ratio of delegates at the meeting of country j. This list of variables reflects the observable component E_j of R_j for regulation i in Section 4.3.2. Additionally, m_{ij} includes the size of the politic opposition in country j (number of veto players) and two measures of trade levels inside and outside Mercosur in the economic sector that the regulation addresses. This completes the list of covariates included in the most basic model that we consider (*Model I* in Table 4.5).

In *Model II*, a similar set of variables Δ_{ijt} is added, controlling for the change in the above list of variables m_{ij} over time compared to their level at the time of contract signing.[16] The idea is to capture the potential change in preferences of the political actors induced by a change in the politic or economic environment compared to the situation when they first signed the regulation.

Note that the political or institutional environment at the time of contract signing may somewhat reflect the environment during the subsequent ratification period. For example,

[16]In addition we control for a change in the political leader of a country compared to the time of contracting.

if public support and media coverage of Mercosur at the time of contract signing is high, it is more likely that this environment will be similar during the ratification period. The potential reputation loss r_j of political actors may be higher during the ratification period, making them more inclined to ratify quickly. Consequently, in *Model I* we expect a positive bias in the estimates of the variables measuring the environment at contract signing. In order to avoid this bias, we control for the change of these variables Δ_{ijt} during the ratification period compared to their level at the time of contract signing in *Models II-III*. *Model III* is motivated by the strong heterogeneity in the ratification durations across the regulations signed in Mercosur. Despite our dataset enabling us to observe many regulation characteristics driving this heterogeneity in ratification hazards, we are most likely unable to explain all variation across regulations by observable variables. Although this problem is usually of no greater concern, due to the duration characteristic of our dataset, the existence of unobserved heterogeneity across regulations can lead to a bias in the estimated baseline hazard and covariate effects (Lancaster, 1990; Van den Berg, 2001) [17]. We expect a bias of the covariate effects towards zero in *Model II* compared to *Model III* (see Tabel 4.5). The bias is avoided in *Model III*, where we account for the effect of unobserved regulation characteristics by a random term V_i.

$$\theta(t|x_{ijt}, V_i) = \lambda(t)\ e^{x'_{ijt}\beta}\ V_i \qquad (\textit{Model III}) \tag{3.6}$$

We assume that V_i is drawn from a Gamma distribution with mean 1 and variance parameter σ^2. Estimates of *Model III* are reported in Appendix 4.A, specifying an inverse Gaussian distribution. The estimated covariate effects are affected little by this change, while the overall model fit is higher in the Gamma model. Note that V_i generates dependence between ratification durations of the four member countries for each regulation. We argue that, conditional on covariates, unobservable regulation characteristics represent the main source of dependence in our dataset.

Other sources of dependence are also possible. For example, if pressure at the negotiation table is relatively high compared to the expected costs of non ratification, actors have an incentive to lie. The theoretical model in Section 4.3.2 predicts that actors secretly opposing the policy change during the ratification period will respond reversely to the

[17] A bias occurs if the underlying true model is the mixed proportional hazard model.

strategy of the international partner as a best response. If player B ratifies quickly, player A will choose to defect if he is of type L. If, on the other hand, player B defects, A will free-ride on the defection efforts of his partner and choose to ratify quickly to avoid reputation costs. Thus, depending on the level of pressure during negotiations, there may exist a negative dependence among ratification durations, caused by strategic interaction. The existence of such an effect could be investigated by using a correlated frailty model in order to estimate the sign of correlation depending on the level of pressure during negotiations. However, this extension is left to further research at this point.

In Table 4.5, we report estimates of the baseline specification without time-varying controls (*Model I*) and the specification with time-varying controls (*Model II*), as well as the mixed proportional hazard model(*Model III*). The first two proportional hazard models rely on the assumption of no dependence within regulations conditional on covariates. On the other hand, the mixed proportional hazard model accounts for a dependence based on unobserved regulation characteristics. It is not a priory clear which model is more suitable to fit the data. In *Model II*, we effectively assume that we observe all relevant regulation characteristics influencing the ratification durations of the four member countries. Conversely, *Model III* is based on the assumption that unobservable influences enter the ratification hazard multiplicatively and are independent of covariates. For this reason, we report both models *I* and *II* and restrict attention to effects robust across the two specifications.

4.5.3 Empirical analysis

We use data on all 1,024 regulations signed between 1994 and 2008 that require ratification at the domestic level in at least one member country of Mercosur.[18] This leads to 3,560 observed ratification durations (Table 4.5). *Model I-III* represent estimations in continuous time; however, the set of time-varying variables Δ_{ijt} and the baseline hazard vary on a yearly basis, leading to 13,057 regulation $\times$ country $\times$ year observations.

In Table 4.5, the estimates of the covariate effects β in equation (5) are reported for *Models I-II* and (6) , with standard errors in parentheses. Note that the estimates of the piecewise constant baseline hazard function and the calendar year dummys are omitted from the estimation outputs. We present point estimates and corresponding confidence

[18]Most regulations need to be ratified by all four members.

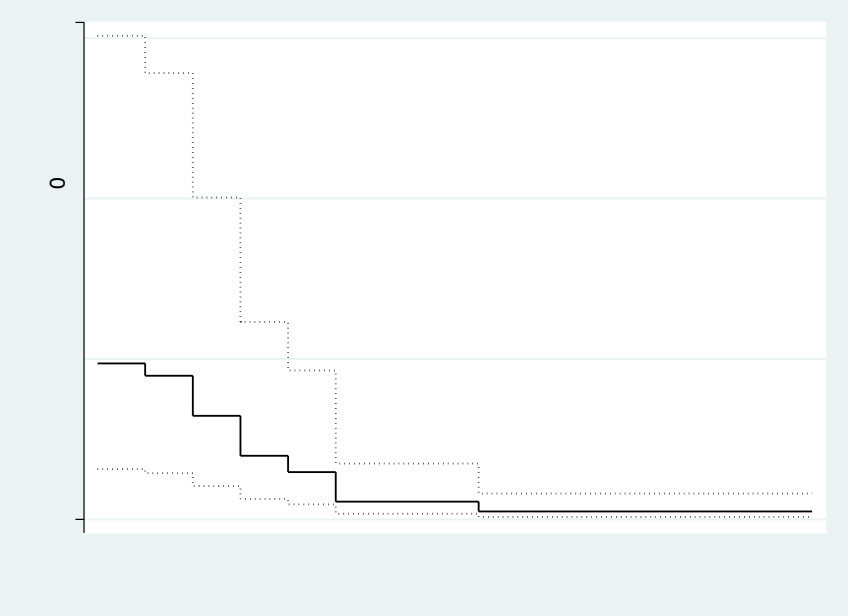
Table 4.5: Parameter estimates of three proportional hazard models with Gamma unobserved heterogeneity distribution

Three Piecewise Constant Models:	**Model I**		**Model II**		**Model III**	
Variable	Estimate	St.Error	Estimate	St.Error	Estimate	St.Error
Member Countries						
Brazil	.235***	(.071)	.118	(.077)	.217***	(.083)
Paraguay	-.095	(.067)	-.149**	(.074)	-.107	(.079)
Uruguay	-.078	(.067)	-.064	(.068)	.044	(.073)
Regulation Characteristics						
Governmental cooperation	-.76***	(.065)	-.775***	(.065)	-1.137***	(.109)
Internal market	-.639***	(.138)	-.641***	(.139)	-1.005***	(.229)
Technical regulations	-.799***	(.061)	-.824***	(.063)	-1.307***	(.099)
Exception to common external tariff	.351***	(.106)	.504***	(.107)	.745***	(.165)
Mercosur interna	-.083	(.131)	-.071	(.132)	-.215	(.213)
Word count	-.026***	(.008)	-.023***	(.008)	-.022**	(.01)
Number articles	-.053	(.164)	-.056	(.165)	-.144	(.255)
Number of references in preamble	.071***	(.011)	.063***	(.011)	.056***	(.018)
Technical annex	.117***	(.048)	.126***	(.048)	.117	(.078)
Overrules Mercosur regulation	-.132**	(.066)	-.125*	(.066)	-.102	(.1)
Year of contract signing	-.07***	(.008)	-.249***	(.024)	-.281***	(.025)
Deadline stated in regulation	.276***	(.058)	.256***	(.06)	.27***	(.098)
Environment at Contract Signing						
Common Market Council	-.656***	(.072)	-.734***	(.073)	-.879***	(.111)
Public support of Mercosur in country	-.009**	(.004)	-.015***	(.005)	-.016***	(.005)
Mercosur presidency	-.123***	(.045)	-.24***	(.064)	-.253***	(.065)
Conflict level in Mercosur	-.103***	(.018)	-.236***	(.055)	-.168*	(.086)
Percentage delegates other countries	-.01***	(.002)	-.01***	(.002)	-.013***	(.002)
Size of opposition in country	-.033*	(.02)	.023	(.024)	.037	(.026)
Exports to Mercosur	-.081*	(.046)	-.07	(.047)	.019	(.054)
Imports from rest of world	.014	(.01)	.025***	(.01)	.039***	(.011)
Change since Contract Signing						
d Public support of Merc. in country	.	(.)	-.006	(.004)	-.005	(.004)
d Presidency of Mercosur	.	(.)	-.087**	(.041)	-.074*	(.042)
d Number of veto players	.	(.)	.074***	(.023)	.108***	(.024)
d Conflict level in Mercosur	.	(.)	-.027**	(.013)	-.012	(.02)
d Exports to Mercosur	.	(.)	-.003	(.063)	.086	(.073)
d Imports from rest of world	.	(.)	-.102***	(.039)	-.104*	(.058)
d Political leader(presidency)	.	(.)	-.15**	(.076)	-.109	(.081)
Unobs. Heterog. (Gamma Dist.)						
log(Variance parameter)	.	(.)	.	(.)	-.706***	(.089)
Variance Parameter					.494	
Log Likelihood	-5313		-5170		-4995	
Regulations	1024		1024		1024	
Regulations × countries	3560		3560		3560	
Regulations × countries × years	13057		13057		13057	

Note: Parameter estimates of three proportional hazard models with a piecewise constant baseline hazard. Estimates are reported as β coefficients of Models *I-III*. Standard errors are reported in parentheses. Estimates with *, ** or *** reflect a 0.1, 0.05 or 0.01 significance level.

bands of the piecewiese constant baseline hazard in Figure 4.6. [19]

Figure 4.6: Estimates of a piecewise constant baseline hazard function



Note: Estimates of the piecewise constant baseline hazard function $\lambda(t)$ of *Model III* in Table 4.5. Dotted lines denote 95% confidence bands.

The main focus of the empirical analysis is to measure the effect of the public, political and institutional environment at the time of contract signing on subsequent ratification durations. The estimates of *Models I-III* suggest that if public and political pressure on international actors during the time of contracting is high, the subsequent ratification process at the domestic level is slowed significantly in the respective country. The estimated effects vary in size yet do not change signs across *Models I-III* (see first five estimates under 'Environment at Contract Signing' in Table 4.5.

In particular, regulations introduced in the Common Market Council have a $1 - e^{-.879} \approx 0.58\%$ (*Model III*) lower ratification hazard compared to those introduced in the Common Market Group or Trade Commission ($1 - e^{-.734} \approx 0.52\%$ in *Model II*). Whereas in the Common Market Group and Trade Commission bureaucrats sit at the negotiation table, in the Common Market Council negotiations are conducted by politicians. In general,

[19] The wider confidence bands for earlier years compared to later years come from rescaling the confidence bands of estimated coefficients β_0 to confidence bands of hazard ratios: confband$(exp(\beta_0)) = [exp(\beta_0^{lower}), exp(\beta_0^{upper})]$. Here, β_0^{lower} and β_0^{upper} denote the confidence intervals of the estimated coefficients β_0.

regulations signed in the Common Market Council receive higher media attention in the four member countries compared to those signed by bureaucrats. The significant negative effect on the ratification hazards suggests that ratification takes significantly longer for regulations that are initially signed by politicians compared to bureaucrats.

In addition, we find a significant negative effect on the ratification hazard if the public is more supportive of the Southern Common Market at the time of contract signing. International actors secretly opposed to a regulation are more inclined to sign it when public opinion in their country is strongly in favor of regional integration. Thus, in terms of ratification at the domestic level, the political actors have no incentive to speed up the process, resulting in a low ratification hazard. A 1% increase in the ratio of supporters in a country lowers the ratification hazard by $1-e^{-.016} \approx 1.6\%$ in *Model III* ($1-e^{-.015} \approx 1.5\%$ in *Model II*).

We find a similar effect for a country holding the presidency of Mercosur at the time of contract signing. The presidency rotates between members, with the main negotiations hosted by the respective country, resulting in higher media coverage of the negotiation outcomes and topics relating to Mercosur in general. This environment puts additional pressure on political actors to signal an attitude towards international cooperation, thus making them more inclined to sign regulations that they have little incentive to ratify later. The estimates in Table 4.5 imply that if a country holds the presidency at the time of contract signing, this corresponds to a $1-e^{-.253} \approx 22.4\%$ (*Model III*) decrease in the subsequent ratification hazard ($1-e^{-.24} \approx 21.3\%$ in *Model II*).

Furthermore, negotiation outcomes are influenced by the overall level of conflict between the four members of Mercosur. In times of high levels of conflict, signals are particularly necessary to communicate a sustained interest in regional cooperation. This places contracting in Mercosur at center stage, given that it is capable of reassuring the partners of an interest in cooperation despite few, but salient conflicts. We find a significant negative effect of $\beta = -.168$ for a 1 point increase in the measure of conflict level constructed by Gómez-Mera (2009), implying a $1-e^{-.168} \approx 15.5\%$ decrease in the subsequent ratification hazard ($1-e^{-.236} \approx 21.0\%$ in *Model II*).

The ratio of delegates at the negotiation meetings representing the other countries may influence whether political actors feel pressured to sign a regulation. The estimates in *Model III* suggest that a 1% increase in the ratio of foreign delegates at the negotiation

meeting lowers the ratification hazard by $1 - e^{-.013} \approx 1.3\%$ in *Model III* ($1 - e^{-.01} \approx 1.0\%$ in *Model II*). [20]

We do not find evidence of an effect for the size of the politic opposition in a country. We also include export levels to Mercosur for the regulations addressing internal trade, as well as import levels from the rest of the world for regulations addressing external trade in the economic sector that the regulation addresses, with these two variables reflecting the economic environment that decision makers face at the time of contract signing. However, we do not find robust evidence for an effect significantly different from zero for *Models II-III*.

Finally, we find strong heterogeneity in ratification hazards across policy types. In particular, we find that regulations addressing governmental cooperation in Mercosur have relatively low ratification hazards compared to the reference category of the common external tariff ($1 - e^{-1.137} \approx 68.0\%$ less in *Model III* and $1 - e^{-0.775} \approx 54.0\%$ less in *Model II*). Political actors find regulations on general governmental cooperation that simply announce future cooperation particularly useful to create a positive image in the press. These regulations allow for making broad cooperative claims without necessarily changing the status quo through costly adaptation. If political actors wish to create a positive image as proactive regional leaders, they should be particularly prone to doing so by relying on these type of policies.

Conversely, regulations addressing the common external tariff are beneficial to all members of Mercosur in most cases . Consequently, ratification is relatively quick. Recall that in the standard case of regional economic cooperation, the common external tariff is one of the most common policy areas for free riding. A country benefits tremendously if all others implement a common, higher tariff, given that private actors prefer to trade with lower tariff barriers. While regional cooperation with short term incentives for defection lead to expect deviations from a common cooperative course, the ratification rules implemented in Mercosur prevent such behavior. Here, empirical results are in line with the theoretical expectations, namely that exceptions from the common external tariff are beneficial policies. Actors are always eager to contract and subsequently ratify such regulations. An

[20] Note, that at each meeting several regulations are signed. Consequently, delegates of all four member countries are present at each meeting but depending on the meeting and the decision body (Common Market Council, Common Market Group or Trade Commission), a different ratio of country delegates is present.

additional benefit from contracting alone does not have any influence on the propensity to put a policy into practice. Considering the short ratification durations for regulations addressing the common external tariff, incentives for free riding clearly do not cause Mercosur's eminent ratification issues.

4.6 Conclusion

This paper shows that cooperation problems that arise from incentives for free riding are not necessarily as intractable as the literature in international relations suggests. So far, implicit and explicit means of retaliation are believed to be the only means of solving such cooperation issues. We argue that rules for ratification may do so just the like. When the last country determines the overall effectiveness of a policy, incentives for individual defection do not arise. However, the second insight of this paper reveals that altering ratification rules in this way does not solve all cooperation issues. If incentives for contracting are high and costs from non ratification low, actors may adopt policies that provide little benefit from collective action. Under these conditions, inefficient contracting may lead to many international rules that are signed, yet not ratified.

We offer empirical evidence for our theoretical considerations, based upon the Southern Common Market, a regional cooperation scheme between Argentina, Brazil, Paraguay and Uruguay, which adopted 1,024 regulations at the regional level between 1994 and 2008. Mercosur is particularly suitable for illustrating our theoretical considerations. First, it uses the above-mentioned rules for ratification, with policies only becoming effective in the whole region once the last country incorporates a regulation into domestic law. Second, due to the general institutional provisions, reputation costs of postponing ratification are particularly low.

We collect and analyze the complete ratification record on all policies and find proof for our theoretical claims. Rules on the common external tariff typically offer individual incentives for defection. By contrast, actors in the Southern Common Market prefer to realize the collective good and do not depart from commonly agreed policies. Against the backdrop of low costs for non ratification, incentives from contracting allures actors to first contract and then refrain from ratification when the utility from collective goods are low. Politicians are eager to use a high popularity of Mercosur to create a positive image among their electorate. In the meantime, ratification rules easily allow for halting

the overall effectiveness of rules that are of little benefit. When bureaucrats decide such effects do not arise, since these decision makers do not share the same incentives as their politically accountable principals. In a similar vein, if politicians adopt policies that create collective goods with a high utility, they do not halt ratification processes and ratification is swift.

This paper makes an important contribution towards understanding how international institutions produce outcomes in international cooperation (Jacobson, 2000; Simmons, 2000; Simmons and Danner, 2010; Von Stein, 2005). The ratification rules that the Southern Common Market uses inhibit free riding, whereby actors incorporate regulations beneficial to all without regress. However, the institutional design encourages inefficient contracting, since actors may adopt policies that provide benefits from contracting only and refrain from their incorporation.

4.A Appendix

Table 4.6: Parameter estimates of three proportional hazard models with Inverse Gaussian unobserved heterogeneity distribution

Three Piecewise Constant Models:	**Model I**		**Model II**		**Model III**	
Variable	Estimate	St.Error	Estimate	St.Error	Estimate	St.Error
Member Countries						
Brazil	.235***	(.071)	.118	(.077)	.198***	(.082)
Paraguay	-.095	(.067)	-.149**	(.074)	-.119	(.079)
Uruguay	-.078	(.067)	-.064	(.068)	.028	(.072)
Regulation Characteristics						
Governmental cooperation	-.76***	(.065)	-.775***	(.065)	-1.045***	(.103)
Internal market	-.639***	(.138)	-.641***	(.139)	-.878***	(.223)
Technical regulations	-.799***	(.061)	-.824***	(.063)	-1.162***	(.095)
Exception to common external tariff	.351***	(.106)	.504***	(.107)	.673***	(.152)
Mercosur interna	-.083	(.131)	-.071	(.132)	-.095	(.2)
Word count	-.026***	(.008)	-.023***	(.008)	-.025**	(.011)
Number articles	-.053	(.164)	-.056	(.165)	-.112	(.255)
Number of references in preamble	.071***	(.011)	.063***	(.011)	.066***	(.018)
Technical annex	.117***	(.048)	.126***	(.048)	.122	(.075)
Overrules Mercosur regulation	-.132**	(.066)	-.125*	(.066)	-.116	(.098)
Year of contract signing	-.07***	(.008)	-.249***	(.024)	-.272***	(.026)
Deadline stated in regulation	.276***	(.058)	.256***	(.06)	.311***	(.094)
Environment at Contract Signing						
Common Market Council	-.656***	(.072)	-.734***	(.073)	-.872***	(.106)
Public support of Mercosur in country	-.009**	(.004)	-.015***	(.005)	-.016***	(.005)
Mercosur presidency	-.123***	(.045)	-.24***	(.064)	-.249***	(.065)
Conflict level in Mercosur	-.103***	(.018)	-.236***	(.055)	-.229***	(.084)
Percentage delegates other countries	-.01***	(.002)	-.01***	(.002)	-.013***	(.002)
Size of opposition in country	-.033*	(.02)	.023	(.024)	.037	(.025)
Exports to Mercosur	-.081*	(.046)	-.07	(.047)	.008	(.053)
Imports from rest of world	.014	(.01)	.025***	(.01)	.035***	(.011)
Change since Contract Signing						
d Public support of Merc. in country	.	(.)	-.006	(.004)	-.005	(.004)
d Presidency of Mercosur	.	(.)	-.087**	(.041)	-.075*	(.042)
d Number of veto players	.	(.)	.074***	(.023)	.103***	(.024)
d Conflict level in Mercosur	.	(.)	-.027**	(.013)	-.022	(.02)
d Exports to Mercosur	.	(.)	-.003	(.063)	.05	(.069)
d Imports from rest of world	.	(.)	-.102***	(.039)	-.104**	(.05)
d Political leader(presidency)	.	(.)	-.15**	(.076)	-.106	(.08)
Unobs. Heterog. (Inv. Gaussian)						
log(Variance parameter)	.	(.)	.	(.)	-.473***	(.125)
Variance Parameter					.623	
Log Likelihood	-5313		-5170		-5022	
Regulations	1024		1024		1024	
Regulations × countries	3560		3560		3560	
Regulations × countries × years	13057		13057		13057	

Note: Parameter estimates of three proportional hazard models with a piecewise constant baseline hazard. Estimates are reported as β coefficients of Models *I-III*. Standard errors are reported in parentheses. Estimates with *, ** or *** reflect a 0.1, 0.05 or 0.01 significance level.

Chapter 5

Inference for Shared-Frailty Survival Models with Left-Truncated Data[1]

5.1 Introduction

In this paper, we consider inference for shared-frailty survival models. These are Mixed Proportional Hazard (MPH) models in which systematic unobserved determinants of duration outcomes are identical within units or groups of individuals. We allow the spell durations to be subject to left-truncation, meaning that the duration outcome is only observed if it exceeds a certain threshold value, and we focus on random-effects likelihood-based inference. We show that the Stata software package command to estimate shared-frailty survival models in the presence of left-truncated duration data should not be applied, given that it maximizes a likelihood function that does not sufficiently take account of dynamic selection before the truncation points.

In order to explain this and motivate the relevance of our contribution, we start with an introduction of the survival models with unobserved heterogeneity (or frailty terms) that are included in Stata for statistical inference. Shared-frailty models are an important class of such models.

Empirical survival studies or studies in duration analysis commonly adopt some version of the Mixed Proportional Hazard (MPH) model for the hazard rate. The MPH model stipulates that the individual hazard rate (or exit rate out of the current state) θ depends on the elapsed duration t, explanatory variables x and unobserved determinants v, such

[1] This chapter is joint work with Gerard van den Berg.

that

$$\theta(t|x,v) = \lambda(t)\phi(x)v$$

at all t, x, v for some functions λ and ϕ (see Lancaster, 1990; Van den Berg, 2001, for surveys). Here, ϕ is the function of interest, although λ is sometimes also of interest. Typically, at least some elements of the vector x are time-varying, although we ignore this in this paper for ease of exposition. Notice that without loss of generality, v can be seen as the joint multiplicative effect of a vector of unobserved determinants on the individual hazard rate. The term v is often called the frailty term. It is not directly estimated from the data, given that it varies across individuals. Moreover, in contrast to linear regression analysis, ignoring unobserved heterogeneity leads to biased estimates of λ and ϕ, because individuals with a high v leave the state of interest on average earlier than individuals with low v. This phenomenon is called "weeding out" or "sorting", which may occur at different speeds for different x, causing the composition of survivors in terms of v to change over time. In general, ignoring this leads to a negative bias in the estimate of $\lambda(t)$ and a bias in the estimated covariate effects (Lancaster, 1990; Van den Berg, 2001). The most common approach for inference is to assume that v has a distribution G in the population and to estimate its parameters along with (the parameters of) λ and ϕ using Maximum Likelihood Estimation, where the likelihood contribution of an individual spell integrates over G. In econometrics, this is called random-effects estimation. To ensure that identification is not fully driven by functional form assumptions, it is assumed that x and v are independently distributed in the population and that $\mathrm{E}(v) = 1$. The population constitutes the inflow into the state of interest (although this may be modified; see below). By far the most common functional form for G is the gamma distribution, which can be justified as an approximation to a wide class of frailty distributions (Abbring and van den Berg, 2007). The approximation improves with left-truncation of the durations. An alternative frailty distribution is the inverse-Gaussian distribution.

It is often natural to assume that small subsets of different individuals or spell durations share the same value of v. For example, different unemployment spells of the same person may share the same unobserved determinant v, or the mortality rates of identical twins may be assumed to depend on identical unobserved determinants v. In general, the data may identify groups or units or strata such that different spells within a group or unit or stratum share the same v. Data with this feature are often called multi-spell duration

data. To keep the terminology simple, consider the case where we observe at most two spells for each unit in the sample. The unit has a given value of v, and we assume that its spell durations are independent drawings from the univariate duration distribution of t given x, v, where, naturally, v is unobserved, so that the durations given x are not independent. Whether x is also identical across spells or individuals within a unit depends on the context. For ease of exposition, we take the data to consist of a random sample of units. We return to this below.

The multi-spell MPH model was first proposed by Clayton (1978) and is nowadays known under the name "shared-frailty model". Notice that it has the same unknown functions as the single-spell MPH model, namely λ, ϕ and G. The empirical analysis of shared-frailty models is widespread (see e.g. Hougaard, 2000, and Van den Berg, 2001, for surveys). If the underlying modeling assumptions are correct, multi-spell data enable identification of the MPH model under weaker assumptions than single-spell data, and the estimation results are more robust with respect to functional-form assumptions (see Hougard, 2000; Van den Berg, 2001, for surveys). By straightforward extension of the estimation with single-spell data, the most common estimation methods are random-effect procedures where each unit or group provides a likelihood contribution that integrates over the distribution G of v across the units and where λ, ϕ and G are parameterized.[2]

The Stata software package offers a large number of pre-programmed estimation routines for survival analysis. In this sense, Stata is unique among the available software packages covering survival analysis; indeed, it has become popular among empirical researchers. The two main survival model estimation commands `streg` and `stcox` also capture the shared-frailty model by invoking the option `shared()` to indicate which individuals share the same value of v. Gutierrez (2002) provides an overview of parametric shared-frailty models in Stata. See Hirsch and Wienke (2012) for an overview of software packages with estimation routines for shared-frailty models.

Sampling schemes where durations are left-truncated are common in both single-spell and in multi-spell survival analysis Guo (1992). For example, unemployment duration spells are often only recorded in register data if the duration exceeds one month. Population

[2]If different individuals within a unit or group have different values of x then Stratified Partial Likelihood Estimation can be used as an alternative (fixed effects) method Ridder and Tunali (1999); Chamberlain (1985); Kalbfleisch and Prentice (2011). In Section 5.3 we give a brief overview of the use of this method in Stata.

register data typically follow individuals from a given point in calendar time onwards, with the starting points of the spells that are ongoing at the beginning of the register's observation window often also observed. The spells that started, for instance, t_0 time units before the beginning of the observation window are subsequently only observed if the duration exceeds t_0. With the increasing availability of such register data in socio-economic and health research, the usage of left-truncated duration data has increased. This also applies to multi-spell data. For example, death causes of Danish twins were only systematically recorded as of January 1, 1943; therefore, in studying death causes among those born before 1943, it makes sense to restrict attention to both twin members being alive on January 1, 1943.[3] If the duration from birth until death due to a specific death cause is the relevant duration variable, then this variable is left-truncated at the age attained on January 1, 1943. Hence, the left-truncation points as measured in the age dimension differ across twin pairs. In studies with hospital patients, only those who survive up to the point when the trial period at the hospital starts are observed. If the patient subsequently experiences remission and relapse, subsequent illness spells may not be left-truncated.

Stata allows for left-truncation of the duration data through the `enter()` option when declaring the data as duration data by the `stset` command. Importantly, the value t_0 of the truncation threshold may differ across individuals (as well as across spells for a given unit, in the case of the shared-frailty model).

Notice that left-truncation gives rise to a second selection issue, in addition to the selection generated by the dynamic weeding-out. After all, surviving up to some threshold value is more likely if the frailty term is small. The Stata routine for shared-frailty models[4] ignores the fact that the second selection impacts on the first selection. Restricting the outcome to exceed a lower threshold implies that the frailty distribution in the sample systematically differs from that in the population upon inflow into the state of interest.[5] If the former distribution is nevertheless assumed to equal the latter, then, as we shall see, the resulting estimators of β and λ are inconsistent. One may redefine the population

[3] After all, if a twin member is observed to have died before 1943 then it is not known whether this was due to the cause of interest or due to another cause. In the latter case, the moment of death due to the cause of interest is right-censored by an event with an unknown distribution, and inference would include the estimation of this distribution.

[4] This routine is available since Version 7, up to and including the current Version 13.

[5] See Ridder (1984) for an account of the differences between frailty distributions in different types of single-spell samples.

to be the survivors at t_0, although this only makes sense if t_0 is identical across all units and spells.

The interplay between left-truncation and dynamic selection has always been recognized in the single-spell survival analysis literature. As we discuss below, the role of this interplay has been obscured with multiple spells. However, we are not the first to point out the importance of dealing with the above interplay, including its implication for the frailty distribution in the sample in shared frailty models. Indeed, Jensen et al. (2004) provide a lucid account. They contrast the correct likelihood function to the likelihood function where the interplay is ignored for the case of gamma-distributed frailties, and they discuss the bias when using the latter. They point out that Nielsen et al. (1992), a seminal paper in survival analysis, used the incorrect likelihood in the case of left-truncated data in the shared frailty model. Elsewhere within the literature, Rondeau and Gonzalez (2005) use the correct likelihood for their semi-parametric estimator of the shared frailty model in the case of left-truncated data, whereas Do and Ma (2010) use the other likelihood function for their semi-parametric estimator in the same setting.

With an update to release 12, that carries over to the current release 13, Stata has reacted to our work. As of this update, estimation of a shared frailty model with `streg` or `stcox` with left-truncated data or gaps generates an error message for the user. The error message explains how both commands `streg` and `stcox` implicitly assume that the corresponding frailty distribution is independent of the covariates and the truncation points. Although it is not recommended to use the commands in this setting, the error message may be overwritten by using the option `forceshared`.

The remainder of the paper is structured as follows. In Section 5.2, we discuss left-truncation in multiple spell duration data in more detail. We show the conditions under which the likelihood function of the parametric model in the `streg,shared()` command is misspecified for left-truncated data, and present the correct likelihood function. In Section 5.3, we demonstrate in a short simulation study with the `streg` command how the magnitude of the bias resulting from the misspecification depends on the level of truncation and variance of the frailty distribution. Additionally, in Section 5.4, we discuss the analogous problem with the `stcox` command in Stata for the semi-parametric estimation of the shared gamma frailty model and discuss how the misspecification may be fixed. We list published articles that use this Stata command to semi-parametrically

estimate the shared gamma frailty model with left-truncated data. Finally, Section 5.5 concludes. In Appendix 5.A.4, we introduce a corrected parametric Stata command called `stregshared`.

5.2 Likelihood specification with left-truncated duration data and shared frailties

Consider a random sample of single spells, where the MPH model applies. The random sample consists of independent draws from the distribution of $T|X$ for various values x of X, where T denotes the random duration variable. We consider likelihood-based inference, and for the moment take λ, ϕ and G to be parametric functions. The spell durations may be independently right-censored, although we are not concerned with that here. Consequently, the likelihood contribution of a single spell is the probability density function $f_u(t|x)$ of $T|X$ evaluated at the observation (t, x), with

$$f_u(t|x) = E_v(f_c(t|x, v)) = \int_v \lambda(t)\phi(x)v \exp(-\Lambda(t)\phi(x)v)dG(v)$$

in which $\Lambda(t) := \int_0^t \lambda(u)du$ denotes the so-called integrated baseline hazard, and f_c is the probability density function of $T|X, V$.

Next, consider a random sample of units, each with $j = 1, 2$ spells that share their frailty term v. Throughout the paper, we assume that, conditional on v and x, the spells are independent. The likelihood contribution of a unit with non-truncated uncensored duration outcomes $t_1|x_1$ and $t_2|x_2$ subsequently equals $\int_v f_c(t_1|x_1, v)f_c(t_2|x_2, v)dG(v)$.

Left-truncation of a single-spell duration outcome variable means that the variable is only observed if its value exceeds a lower threshold, say t_0. Throughout the paper, we are only concerned with deterministic t_0. In a random sample of left-truncated single spells, the individual likelihood contribution equals $f_u(t|x)/(1 - F_u(t_0|x))$, with F_u being the distribution function associated with the density f_u. With multiple spells per unit (or group or stratum), left-truncation of a spell duration outcome can be defined analogously, regardless of whether other spells are observed for this unit where the outcome exceeds its lower threshold. However, sometimes none of the duration outcomes of a unit are observed or used if at least one of them is left-truncated. The study of cause-specific mortality with

twin data mentioned in Section 5.1 is such an example. It is useful to consider this case first for expositional reasons. If the number of spells (observed or not observed) of a unit is known, the model can be used to derive the likelihood function. Suppose that each unit consists of two spells $j = 1, 2$, which are observed conditional on both spell durations surviving up to their truncation points t_{01} and t_{02}, respectively. This might be called "strong left-truncation". In the simple case of no censoring, the likelihood contribution L of the unit is now given by the density function of $t_1, t_2 | T_1 > t_{01}, T_2 > t_{02}, x$, which can be expressed as

$$L = \int_0^\infty f_c(t_1|T_1 > t_{01}, x_1, v) f_c(t_2|T_2 > t_{02}, x_2, v)\, dG(v|T_1 > t_{01}, T_2 > t_{02}, x) \tag{5.1}$$

with $x = (x_1, x_2)$ and T_j denoting the random duration variables. Therefore, we average over the conditional frailty distribution $G(v|T_1 > t_{01}, T_2 > t_{02}, x)$ in units where both spells survive up to their truncation points t_{0j} (and given x). This is the distribution of v in the sample of observed spells, which can be expressed in terms of the model primitives through

$$dG(v|T_1 > t_{01}, T_2 > t_{02}, x) = \frac{(1 - F_c(t_{01}|x_1, v))(1 - F_c(t_{02}|x_2, v))dG(v)}{\int_0^\infty (1 - F_c(t_{01}|x_1, w))(1 - F_c(t_{02}|x_2, w))dG(w)}$$

where

$$1 - F_c(t_{0j}|x_j, v) = \exp(-\Lambda(t_{0j})\phi(x_j)v).$$

Note that even if only one of the spells j within a unit has $t_{0j} > 0$, the distribution $G(v|T_1 > t_{01}, T_2 > t_{02}, x)$ differs from $G(v)$.

Assuming a gamma-distributed frailty with $E(v) = 1$ and $Var(v) = \sigma^2$ yields[6]

$$L = \phi(x_1)\lambda(t_1)\phi(x_2)\lambda(t_2)(\sigma^2 + 1)(1 + \sigma^2 M(t_{01}, t_{02}))^{1/\sigma^2}(1 + \sigma^2 M(t_1, t_2))^{-(1/\sigma^2+2)}, \tag{5.2}$$

where $M(t_1, t_2) = \phi(x_1)\Lambda(t_1) + \phi(x_2)\Lambda(t_2)$. For ease of exposition, note that we omit the dependence of M on x_1, x_2.

Rather than the above type of left-truncation, we may consider sampling schemes with different types of reduced observability of low spell durations in a shared-frailty model.

[6]See Appendix 5.A.1 for details.

If only one spell per unit is not left-truncated, one may nevertheless include it in the data used for inference. However, given that the number of spells per unit equals two, we directly infer that the other spell duration t_j satisfies $t_j \leq t_{0j}$. In other words, t_j is left-censored rather than left-truncated. The unit then provides a likelihood contribution equal to $\int_v f(t_1|x_1, v)F(t_2|x_2, v)dG(v)$, where we took $j = 2$ and F_c denotes the cumulative distribution function of $t_2|x_2, v$.

Alternatively, the number of spells per unit may not be fixed and may increase with the sample size. Jensen et al. (2004) provide a detailed formal likelihood derivation in a rather general dynamic sampling framework where the number of (possibly simultaneously occurring) spells per unit may increase with the time that units are followed, and where all observed spells per unit are used for the statistical inference. Under some assumptions, the likelihood contributions are identical to equation (5.1). In particular, if two spells are observed for some unit, then the distribution of the frailty term of this unit, conditional on the two spell durations exceeding t_{01} and t_{02}, respectively, equals $G(v|T_1 > t_{01}, T_2 > t_{02}, x)$.[7] Equation (5.2) replicates likelihood equations in e.g. Jensen et al. (2004) and Rondeau and Gonzalez (2005) for the shared gamma frailty model with left-truncated data.

We now turn to the likelihood function used in Stata. The Stata Manual (e.g., see Stata, 2009, p.383) provides a likelihood contribution for the case of two possibly left-truncated spells and a shared gamma frailty model. This is used in the `streg` command with the options `frailty(gamma)` and `shared()`. In the absence of right-censoring, the likelihood contribution states that[8]

$$L_{Stata} = \phi(x_1)\lambda(t_1)\phi(x_2)\lambda(t_2)(\sigma^2 + 1)(1 + \sigma^2(M(t_1, t_2) - M(t_{01}, t_{02})))^{-(1/\sigma^2+2)}. \quad (5.3)$$

which evidently differs from Equation (5.2). It is shown in Appendix 5.A.2 that the right

[7] Because of the dynamically evolving sampling scheme, where new spells per unit may start during the observation window, they need to make an approximation to deal with changes in the composition of the inflow during the observation window. This is an additional complication that does not affect the issues we focus on but which does not allow us to draw on their simulation results to assess the bias due to ignoring the interplay between left-truncation and dynamic selection.

[8] We translate the notation of the Stata Manual, as follows: $S_{ij}(t_{ij}) = e^{-\phi(x_{ij})\Lambda(t_{ij})}$ and $h_{ij}(t_{ij}) = \phi(x_{ij})\lambda(t_{ij})$, where we omit the index i.

hand side of Equation (5.3) can be rewritten as

$$L_{Stata} = \int_0^\infty f_c(t_1|T_1 > t_{01}, x_1, v) f_c(t_2|T_2 > t_{02}, x_2, v)\, dG(v) \tag{5.4}$$

where $G(v)$ is a gamma distribution. This expression corresponds to the likelihood contribution presented in (Gutierrez, 2002, p.34) for general frailty distributions. By comparing Equations (5.4) and (5.1), it is clear under which conditions Equations (5.3) and (5.2) differ, as well as the underlying reason for them to differ. First, they differ if and only if $Var(v) > 0$ and at least one of the following inequalities applies also: $t_{01} > 0$, $t_{02} > 0$. Secondly, they differ because the conditional densities in Equation (5.4) are averaged over the inflow distribution $G(v)$ rather than the frailty distribution $G(v|T_1 > t_{01}, T_2 > t_{02}, x)$, conditional on the spell durations being left-truncated. The critical issue is that the likelihood in (5.3) treats the data as if no sorting had taken place prior to the beginning of the observation window. Therefore, it is implicitly assumed that the inflow distribution of frailties at $t = 0$ does not change until the point of truncation. However, since the subjects are at risk from $t = 0$ onwards, this assumption cannot hold. The above problem carries over to the case where the frailty is assumed to follow an inverse-Gaussian frailty distribution in the `streg` command. The likelihood function for a shared frailty model with shared inverse-Gaussian frailties and left-truncated duration data is derived in Appendix 5.A.3. This may be contrasted to the function given in the Stata Manual (Stata, 2009, p.383).

An ad-hoc approach to deal with the discrepancy between the likelihood function and the Stata routine is to simply assume from the outset that the frailty distribution in the sample does not depend on x and the truncation points. This effectively amounts to a redefinition of the population, as the inflow into the state of interest at the moment of left-truncation, with the assumption that in this newly defined population, v is independent of x and of the elapsed time spent in the state of interest at the truncation point. Under this assumption, the Stata likelihood is correct. If the truncation points are not dispersed in the original population then such an approach may make sense. It replaces the assumption that v and x are independent in the inflow into the state of interest with the assumption that they are independent at the moment of truncation. If an MPH model guides the exit rate between the inflow and the truncation point, the latter assumption generally entails that x and v are dependent in the original population that constitutes the inflow into the

state of interest.

However, if the truncation points t_{0j} are dispersed then this approach does not make much sense. For example, consider two units i, i' each with two spells j. The units have identical systematic duration determinants including identical x within and across units, although their left-truncation points differ. In obvious notation, we take $0 < t_{0i1} = t_{0i2} < t_{0i'1} = t_{0i'2} < \infty$, so that there is no dispersion of truncation points within each unit. The ad-hoc approach would require the distribution of v in the first unit at t_{0i1} to equal the distribution of v in the second unit at $t_{0i'1}$. However, in the first unit, in between t_{0i1} and $t_{0i'1}$, the frailty distribution changes with time in accordance to the shared frailty model, leading to a different distribution at $t_{0i'1}$ than at t_{0i1}. By implication, the distributions of v at $t_{0i'1}$ would differ across units, not because the units behave differently, but rather due to the way in which they have been sampled.

The Stata issues discussed thus far refer to the use of the options `shared` and `frailty()` in the `streg` command, in conjunction with the use of the option `enter()` in the command `stset`. The `streg` command with the options `shared` and `frailty()` corresponds to parametric shared-frailty models. However, Stata also offers a routine for the semi-parametric estimation of shared-frailty models, which can also be applied in the case of left-truncated data. In Section 5.4, we discuss this routine in detail and explain how it suffers from a very similar misspecification as in the parametric case.

We finish this section by revisiting the cases where the Stata likelihood function and our own likelihood function coincide. Recall that if none of the spells are left-truncated then they coincide, and if there is no systematic unobserved heterogeneity (so $\text{Var}(v) = 0$) then they coincide as well. If a unit or group always consists of one single spell, the Stata likelihood and our likelihood do not coincide, but our likelihood should then coincide with the likelihood of the MLE estimator for a single-spell MPH setting with left-truncated data. We know that the latter is correctly specified in Stata. By implication, with left-truncated data, the Stata estimator for the shared frailty model with a single spell per unit does not equal the Stata estimator for the corresponding MPH model with single-spell data. This is readily verified. In the latter case, the frailty distribution conditions on survival until the truncation point, whereas in the former case it does not.

According to Hirsch and Wienke (2012), none of the other software packages with estimation routines for shared-frailty models allow for left-truncation, with the exception

of an R package called Frailtypack. This uses the semi-parametric Rondeau and Gonzalez (2005) estimator, which uses a full likelihood function that takes account of the interplay between dynamic selection and left-truncation (their estimator penalizes non-smoothness of the baseline hazard function $\lambda(t)$).

5.3 Simulation results

Recall that we are not primarily interested in small-sample properties of estimators, but rather in the appropriate choice of likelihood function, which should be visible in estimates based on a large sample. We simulate data from a shared frailty model, with the sample consisting of units each comprising two spells with a shared gamma frailty. The baseline hazard $\lambda(t)$ follows either a Weibull specification ($\lambda(t) = \alpha t^{\alpha-1}$) or a Gompertz specification ($\lambda(t) = e^{\alpha t}$). Furthermore, $\phi(X) = e^{X\beta}$ with $X = (1\ x)$ and x being a single time-constant covariate.

In a first step, the covariate x_{ij} is drawn from a standard normal distribution for each spell j of unit i, and the frailty term v_i is drawn from a gamma distribution with $E(v) = 1$ and $Var(v) = \sigma^2$ for each unit i. The unknown model parameters are $\beta \equiv (\beta_0, \beta_1), \alpha$ and σ^2. These have the following possible values:

$$\beta_0 = 0, \quad \beta_1 = 1, \quad \alpha = 1, \quad \sigma^2 \in \{0.5, 1, 2\}. \tag{5.5}$$

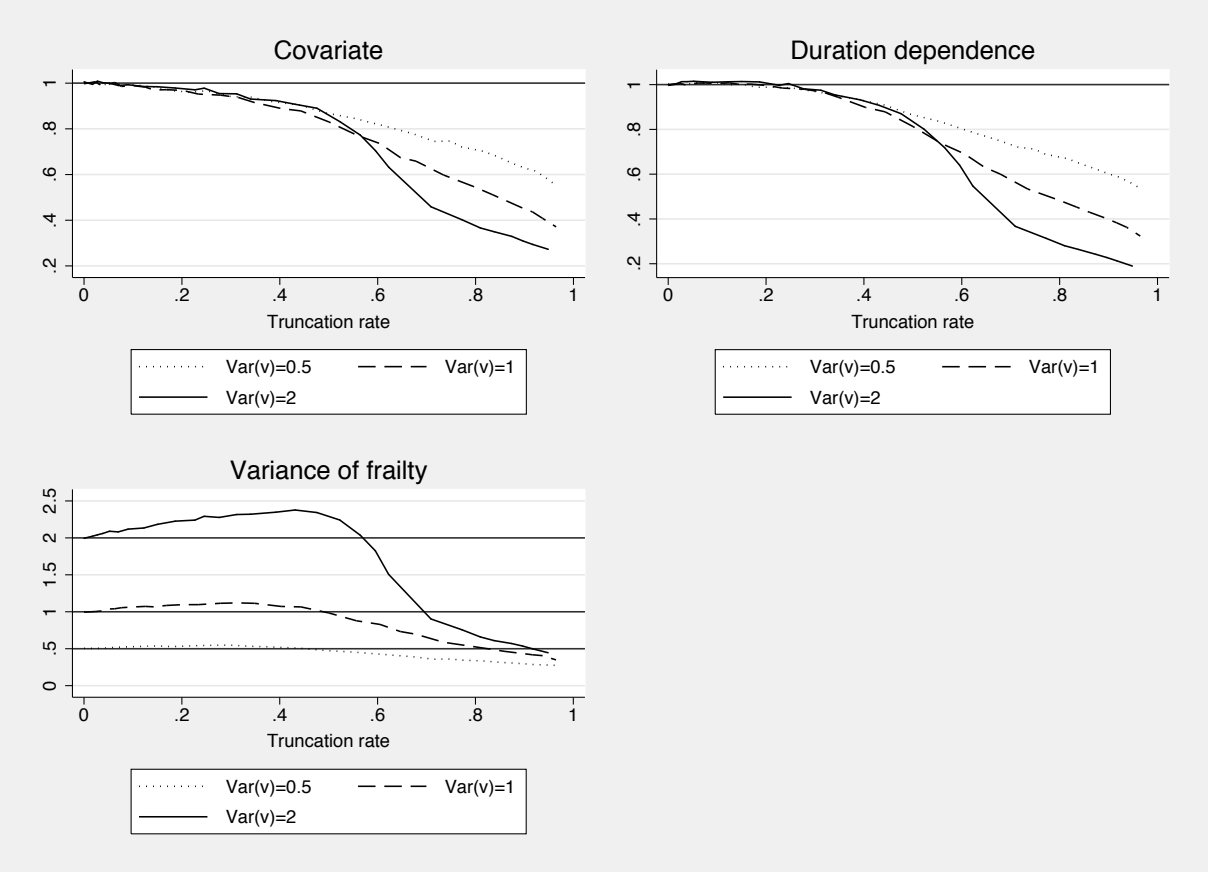
In the case of the Weibull model, $\alpha = 1$ implies the Exponential model with no duration dependence of the baseline hazard, whereas the Gompertz model with $\alpha = 1$ incorporates a strong positive duration dependence. We present simulation results for both cases to reflect applications ranging from economics to mortality studies. Furthermore, we distinguish between three different values of the variance σ^2 of the frailty distribution. These values are in line with those in the simulations in Jensen et al. (2004).

In a second step, for given covariates, frailty terms and parameter values, the durations t_{i1} and t_{i2} are drawn independently from the distributions $F_c(t_j|x_{ij}v_j)$, $j = 1, 2$, respectively.[9] Next, we draw the left-truncation thresholds t_{0i1} and t_{0i2} from a uniform distribution with

[9] We use the following transformation of the variable u drawn from a uniform distribution $U(0,1)$: $t_{ij} = \alpha^{-1}\log(1 - \alpha\log(1-u_{ij})(e^{X_{ij}\beta}v_i)^{-1})$ which is the inverse of the cumulative distribution function $F_c(t_{ij}|X_{ij}, v_i) = 1 - exp(-e^{X_{ij}\beta}\alpha^{-1}(e^{\alpha t_{ij}} - 1)v_i)$.

range $(0, b)$. All units with $t_{i1} \leq t_{0i1}$ or $t_{i2} \leq t_{0i2}$ are dropped. This way the sample only contains those units for which both spell durations exceed their left-truncation points. The fraction $c \in [0, 1]$ of data that are dropped due to left-truncation can be fine-tuned by modifying b. Effectively, the sample size of 50,000 units is determined by the requirement that each of the spells of these units has a duration exceeding a left-truncation point. In fact, if the data are sampled from the Exponential model and if σ^2 is large, the estimation of the parameters β_0, α is numerically cumbersome.[10] This suggests that a larger sample is needed for reliable inference, however in light of the computational burden, we opt for the alternative of assuming that the researcher knows that $\beta_0 = 0$.

Figure 5.1: Simulation results of an Exponential shared gamma frailty model

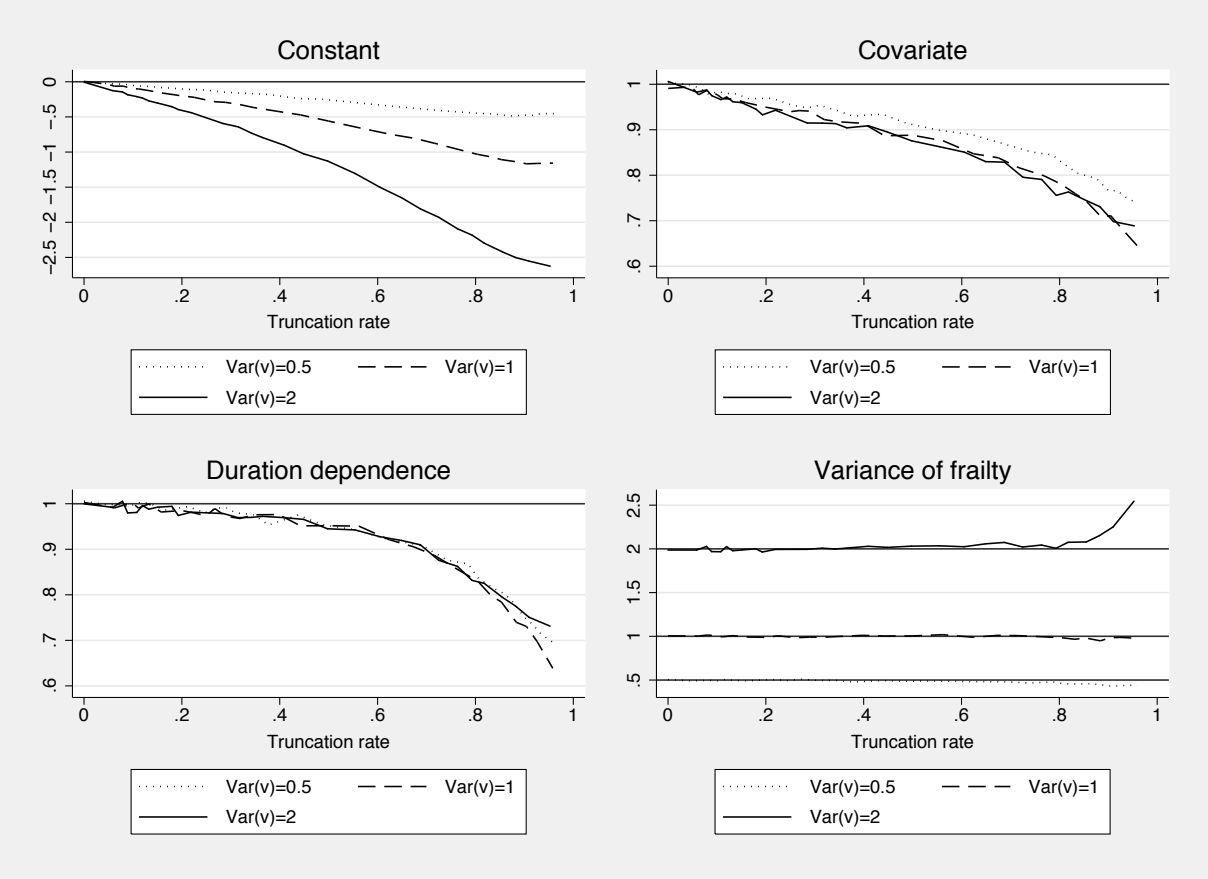


Note: Simulation results of an Exponential shared gamma frailty model with left-truncated data using the Stata command `streg` with the option `shared()`.

In the last step of the simulation procedure we use the `stset` and `streg` commands to

[10] More precisely, the estimation routine suffers from occasional numerical problems. This even occurs in the absence of left-truncation ($c = 0$) if $\sigma^2 \geq 4$.

Figure 5.2: Simulation results of a Gompertz shared gamma frailty model



Note: Simulation results of a shared gamma frailty model with Gompertz duration dependence and left-truncated data using the Stata command `streg` with the option `shared()`.

estimate a shared frailty model in Stata,

```
. stset duration, failure(cens==0) enter(t0)
. streg x , distribution(gompertz) frailty(gamma) shared(id) nohr
```

The results are summarized in Figures 5.1 and 5.2. The panels show the estimates of the constant β_0 (in the case of the Gompertz specification), the covariate effect β_1, the Gompertz duration dependence parameter α, and the variance σ^2 of the gamma frailty distribution. We performed separate simulations with 30 different truncation rates $c \in [0, 1)$, and we connect the resulting points to obtain the displayed curves.

All estimates move away from their true value as the truncation rate c increases from zero. In particular, the covariate effect and the level of the hazard rate are under-estimated at any positive truncation rate.

In general, this is to be expected. As c increases, the simulated distributions of t_{0i1} and t_{0i2} move to the right, and thus the difference between $G(v)$ and $G(v|T_{i1} > t_{0i1}, T_{i2} > t_{0i2}, x)$ increases. With truncated data, units with large v will have exited the state relatively often before having reached the truncation point, so the mean of $v|T_{i1} > t_{0i1}, T_{i2} > t_{0i2}, x$ decreases in t_{0ij}. As the mean of the frailty distribution is fixed to $E(v) = 1$ in the estimation, this decrease in the mean is compensated by an under-estimation of the magnitude of the other determinants of the level of the individual hazard rate (which by themselves have increasing effects on the individual hazard rate).

The bias towards zero of the estimate β_1 can be explained analogously. The true frailty distribution after truncation $G(v|T_{i1} > t_{0i1}, T_{i2} > t_{0i2}, x)$ depends on the covariates x. Spells with a large value of $\exp(X_{ij}\beta)$ as well as a large v_i terminate on average earlier than other spells. Therefore, in the case of a positive β_1, an observation in the truncated sample with a large x is more likely to have a small v_i than observations with a low x. The association between x and the observed hazard rates right after the truncation point is therefore smaller than β_1. If one neglect this by ignoring the dynamic selection before the truncation point, the resulting estimate of β_1 will be biased towards zero.

Figures 5.1 and 5.2 also show that the bias of the estimates depends on the variance of the frailty distribution. As the latter increases, the estimates of the hazard level and the covariate effect move further away from their true values. Again, this is what would be expected. Notice that none of the biases vanish for the sample size $n \to \infty$ for a given truncation rate.

It is important to bear in mind that the simulation results in Figures 5.1 and 5.2 depend on the choice of baseline hazard and the gamma frailty distribution, as well as the choice of the parameter values. For different models, the magnitude of the bias may differ from the presented results.

For Stata users who wish to avoid misspecification of the likelihood function when estimating shared frailty models with left-truncated duration data, we programmed the Stata command `stregshared`, implementing the changes to the likelihood discussed in Section 5.2. In Appendix 5.A.4, we provide a brief description of this new command. Simulations using `stregshared` confirm that the estimator is correct and that the estimates converge to their true values as $n \to \infty$ independent of the level of truncation.

An alternative to the shared frailty model is stratified partial likelihood estimation,

if different individuals within a unit or group have different values of x (Ridder and Tunali, 1999; Chamberlain, 1985; Kalbfleisch and Prentice, 2011). This alternative model is implemented in Stata with the option `strata()` in the `stcox` command. It has the advantage, that the functional form assumptions imposed on the frailty distribution in the shared frailty model, are not needed. However, with this (fixed effects) method only variation within units is exploited and thus a substantial part of the information in the data is lost.

5.4 Inference for semi-parametric shared frailty models

The Stata command `stcox` with the option `shared()` allows for the semi-parametric estimation of a shared-frailty model where $G(v)$ is assumed to be a gamma distribution, $\phi(x) = exp(x'\beta)$, and $\lambda(t)$ is an unspecified function (Cleves et al., 2008). This command can be used in conjunction with the left-truncation option `enter()` in the command `stset`.

5.4.1 Inference with non-truncated data

The semi-parametric estimation method is developed by Therneau and Grambsch (2000), who do not discuss left-truncation of the duration data. The estimator maximizes a profiled penalized partial likelihood (PPL) using two nested loops. In the inner loop, for a given value of the frailty variance parameter θ, the following log penalized partial likelihood function is maximized to derive optimal values for the vector of covariate coefficients β and the frailty vector v

$$
\begin{aligned}
PPL(\beta, v; \theta) &= l(\beta, v) - g(v; \theta) \qquad (5.6)\\
\text{with} \quad l(\beta, v) &= \sum_{j=1}^{U} \sum_{i \in U_j} [(x_i\beta + v_i) - \log \sum_{l \in R_j} \exp(x_l\beta + \log v_l)]\\
\text{and} \quad g(v; \theta) &= -\frac{1}{\theta} \sum_{g=1}^{G} (log(v_g) - v_g).
\end{aligned}
$$

In this penalized model, β reflects the unconstrained and v the constrained parameter vector. The index $j = 1, ..., U$ runs over the ordered failure times, U_j denotes the set of failures at time t_j and R_j is the set of observations k that have not failed before t_j (that is, all k such that $t_j \leq t_k$). The function $l(\beta, v)$ is the log partial likelihood function from a standard Cox regression with optimal values of the frailty terms v treated like observed covariate effects[11]. The penalty function $g(v; \theta)$ in (5.6) reflects the negative log density of the gamma frailty distribution. It penalizes the distance between the fitted gamma distribution and the estimated frailty terms.

In the outer loop, the optimal values for β and v depending on the value of θ are substituted in (5.6), resulting in the log profile penalized partial likelihood

$$PPL(\theta) = \log L(\beta(\theta), v(\theta); \theta). \tag{5.7}$$

Therneau and Grambsch (2000, p.256-258) show that the observed data log-likelihood can be expressed as

$$L(\theta) = PPL(\theta) + \sum_{g=1}^{G} [\frac{1}{\theta} - \log(\frac{1}{\theta} + d_g) - \frac{1}{\theta} \log \theta + \log(\frac{\Gamma(\frac{1}{\theta} + d_g)}{\Gamma \frac{1}{\theta}})]. \tag{5.8}$$

Here, d_g is the number of failures in group g. Therneau and Grambsch further suggest that it is useful to maximize $L(\theta) + \sum_{g=1}^{G} d_g$ rather than (5.8), since the $PPL(\theta)$ converges to $l(\hat{\beta}) - \sum_{g=1}^{G} d_g$ as θ goes to zero. The Stata command `stcox` maximizes $L(\theta) + \sum_{g=1}^{G} d_g$ based on (5.8) over θ and the final estimates of β and v are obtained by maximizing the log penalized partial likelihood in (5.6) using the optimal value of θ.

We should point out that apart from the issues discussed in this paper, the `stcox` command with the `shared` option also has the disadvantage of under-estimating the reported standard errors of the estimated β coefficients, given that they are obtained under the assumption that the true variance of the gamma frailty distribution equals the estimated variance (Cleves et al., 2008).

An alternative to using the penalized partial log-likelihood in (5.6) in the inner loop is to directly maximize the full likelihood using the EM algorithm (see Parner, 1997; Therneau and Grambsch, 2000). The advantage of the PPL over the full likelihood

[11] The expression for $l(\beta, v)$ in Equation (5.6) handles ties by using the Peto-Breslow approximation (Peto, 1972; Breslow, 1974).

approach constitutes that the baseline hazard function λ drops from the partial likelihood in (5.6), thus strongly reducing the parameter space to optimize over. Consequently, this renders estimation using the PPL much faster compared to using the EM approach.
In their description of the PPL, Therneau and Grambsch (2000) do not mention how to deal with left-truncated data. Nevertheless, Stata offers the option to estimated the semi-parametric shared frailty model with left-truncated data. In fact, in Stata only the at risk set R_j in (5.6) is adjusted for truncation in the inner loop, i.e. R_j is adjusted to the set of observations k that have not failed before t_j and have additionally passed their truncation point t_{0k} (that is, all k such that $t_{0k} < t_j \leq t_k$). However, since the penalty function $g(v;\theta)$ in (5.6) that reflects the frailty distribution remains unchanged, the resulting profile penalized partial likelihood suffers from the same misspecification as in the parametric case that we outline in detail in our paper. In the same way as `streg`, the `stcox` command ignores the weeding-out process before the left-truncation points, affecting the distribution of unobserved determinants.
These findings are confirmed by our simulation results using the command `stcox` with the option `shared` in a similar manner as reported for the Weibull or Gompertz model in the previous section. Given that the semi-parametric estimator does not impose the Weibull or Gompertz functional form for the duration dependence λ, standard errors are larger than above. However, point estimates should be close to their asymptotic values with our sample size. Instead, it emerges that the estimates are similar to those obtained with the appropriate `streg` command, for all truncation rates c considered.

5.4.2 Left-truncated data: correcting the misspecification

Correcting the evident misspecification in the semi-parametric estimation command is not as straightforward as in the parametric case. In fact, with left-truncated data, the PPL of Therneau and Grambsch can no longer be used; instead, a full likelihood approach has to be employed. In the following, we explain in detail why the concept of using a penalized partial model does not carry over to left-truncated data, and show how the full likelihood approach using the EM algorithm suggested by Parner (1997) can be adjusted to account for left-truncation in the data. We implement this adjusted estimator. However, simulations reveal that due to the large parameter space to optimize over and the additional uncertainty that comes from the approximation of the baseline hazard at

the truncation points $\lambda(t_0)$, the estimator is highly unstable in samples of reasonable size. We conclude that further research is needed in this area before a usable estimator can be offered.

For non-truncated data, Therneau and Grambsch (2000, p.253-255) provide a justification for using the penalized partial likelihood estimator presented in (5.6) for the case of gamma frailties. They show that for a given frailty variance parameter θ, the solution to the penalized score equations of (5.6) coincides with the solution to the EM-algorithm of Parner (1997), which is based on the full likelihood. However, this result does not carry over to the case of left-truncation. In particular, in order to account for the change in the frailty distribution caused by left-truncation, the penalty function $g(v;\theta)$ in (5.6) would have to be replaced by a function that reflects the negative frailty log density conditional on the durations T having passed their truncation points t_0

$$\begin{aligned} g_{trunc}(v|T > t_0, x; \theta) &= -\log \prod_{g=1}^{G} \frac{(1 - F_c(t_{0g}|x_g, v_g)) f_v(v_g;\theta)}{(1 - F_u(t_{0g}|x_g;\theta))} \\ &= -\sum_{g=1}^{G} \left[\log f(v_g;\theta) - A_{0g}(\lambda,\beta) v_g - \frac{1}{\theta} \log(1 + \theta A_{0g}(\lambda,\beta)) \right] \quad (5.9) \\ \text{with} \quad A_{0g}(\lambda,\beta) &= \sum_{l \in Q_g} \exp(x_l \beta) \Lambda(t_{0l}) \quad \text{and} \quad \Lambda(t) = \int_0^t \lambda(u) du. \end{aligned}$$

Here, the index g runs over the groups $1, ..., G$, f_v is the density function of V, F_u is the cumulative distribution function of $T|X$, F_c the one of $T|X, V$ and Q_g denotes the set of observations in group g. In contrast to the penalty function $g(v;\theta)$ in (5.6), the conditional log density function in (5.9) depends on the baseline hazard function λ as well as the covariate coefficients β. In a penalized partial likelihood such as in (5.6), the partial likelihood function $l(\beta, v)$ may depend on both parameters, constraint (v) and unconstrained (β), while the penalty function $g(v;\theta)$ depends on the constraint parameter alone (see Therneau and Grambsch, 2000, p.120). This central property of a penalized model would be violated if $g(v|T > t_0, x;\theta)$ were to be used as a penalty function. Consequently, with left-truncated data, the resulting full likelihood can not be conveniently reduced to a penalized partial likelihood. Furthermore, the main advantage of the PPL, namely that the baseline hazard function λ drops from the penalized partial likelihood in (5.6), can no longer be exploited. In other words, with left-truncated data, it is necessary to go back to the full likelihood approach in line with Parner (1997), which

we briefly outline in the following.

In the inner loop, maximization of the full likelihood with the EM algorithm involves the two usual steps. In the Maximization step, the function λ and the parameter vector β is estimated using a standard Cox regression, with v treated as a fixed value or offset. In the Expectation step, the conditional expectation of v given the data is computed using the estimates of λ and β from the previous step. In the case of gamma frailties, the computation is straightforward (see Nielsen et al., 1992; Klein, 1992; Therneau and Grambsch, 2000, p.253)

$$E(v_g|T=t,x) = \frac{d_g + \frac{1}{\theta}}{A_g(\lambda,\beta) + \frac{1}{\theta}}. \tag{5.10}$$

Adjusting the EM algorithm in the inner loop to the case of left-truncated data simply involves a small change in the Maximization Step. The set of observations at risk in the partial likelihood is adjusted to account for the left truncation points (that is, $R_j = \{k : t_{0k} < t_j \leq t_k\}$. There is no need to explicitly account for the change in the frailty distribution in the Expectation step, given that the estimates of v are computed using all information in the data (see Equation (5.10)), thereby capturing the effects of left-truncation.

In addition to adjusting the Maximization step of the EM algorithm in the inner loop, we also have to account for the changes in the outer loop in the log profile likelihood. This involves deriving the full data log-likelihood function conditional on the durations having passed their truncation points. The log likelihood contribution for group g is given as (for ease of notation we ignore censoring here)

$$\begin{aligned}
&\log f_{trunc}(t|T>t_0,x) \\
=&\log \frac{\int_v \prod_{l\in Q_g} f_c(t_l|x_l,v_g)\ dG(v)}{\int_v \prod_{l\in Q_g}(1-F_c(t_{0l}|x_l,v_g))\ dG(v)} \\
=&\log \frac{\prod_{l\in Q_g}(e^{x_l\beta}\lambda(t_l))(-1)^{d_g}L_v^{(d_g)}[A_g(\lambda,\beta)]}{L_v[A_{0g}(\lambda,\beta)]} \\
=&\sum_{l\in Q_g}\log(e^{x_l\beta}\lambda(t_l)) - (\frac{1}{\theta}+d_g)\log(1+\theta A_g(\lambda,\beta)) + \frac{1}{\theta}\log(1+\theta A_{0g}(\lambda,\beta)) + m_g(\theta) \\
&\text{with}\ \ m_g(\theta) = d_g\log\theta + \log\frac{\Gamma(\frac{1}{\theta}+d_g)}{\Gamma(\frac{1}{\theta})}\ \ \text{and}\ \ A_g(\lambda,\beta) = \sum_{l\in Q_g} e^{x_l\beta}\Lambda(t_l).
\end{aligned} \tag{5.11}$$

Here, L_v is the Laplace transform with respect to the frailty distribution $G(v)$ and $L_v^{(q)}$

is its q's derivative. The last equality follows from the assumption of gamma-distributed frailties, which implies $L_v(s) = (1+\theta s)^{-\frac{1}{\theta}}$ and $L_v^{(q)}[s] = (-\theta)^q(1+\theta s)^{-(\frac{1}{\theta}+q)}\frac{\Gamma(\frac{1}{\theta}+q)}{\Gamma(\frac{1}{\theta})}$. In the outer loop, the optimal values for $\Lambda(t_l)$, $\Lambda(t_{0l})$ and β from the inner loop are substituted in (5.11) and the resulting profile log-likelihood

$$\log L_{trunc}(\Lambda(t_l;\theta), \Lambda(t_{0l};\theta); \beta(\theta); \theta) \tag{5.12}$$

is maximized over θ. In contrast to the log profile likelihood in (5.7), here the cumulative baseline hazard functions evaluated at each failure time $\Lambda(t_l;\theta)$ and each truncation point $\Lambda(t_{0l};\theta)$ need to be computed based on the current value of θ. Estimates of $\Lambda(t_l;\theta)$ are taken from the solution to the EM algorithm of the inner loop. However, the points $\Lambda(t_{0l};\theta)$ need to be approximated based on the estimates of the cumulative baseline hazard function at the realized failure times $\Lambda(t_l;\theta)$.

Depending on the lack of dispersion in the truncation points across the sample, this approximation can be highly imprecise. Indeed, the extreme case of no dispersion is not an uncommon one. In several applications, every spell in the sample has the same truncation point t_0. For example, the life-spans of twins are sometimes only observed if they have reached a certain age t_0. Furthermore, unemployment spells are often only recorded in register data if the duration exceeds one month. In these examples no failures are observed on the interval 0 to t_0. Nonetheless, in order to account for the weeding-out process over this interval, it is necessary to know the baseline hazard function over the same interval 0 to t_0. In parametric models such as the Gompertz model, the baseline hazard function over the missing years 0 to t_0 is effectively extrapolated from the interval $[t_0, \infty)$ based on the parametric form assumed. In the semi-parametric case, no information on the function $\lambda(t)$ over the interval $[0, t_0)$ exists and thus approximation becomes increasingly arbitrary with longer truncation intervals. In a more favorable truncation scheme, truncation points are evenly spread out on some interval $[0, c)$ and sufficiently many failures are observed over this range.

Simulations with grouped duration data and dispersed truncation points show that estimations even turn out to be very slow and unstable with this favorable truncation scheme, with estimates strongly depending on the choice of starting values. Due to the computational complexity of the EM approach, our simulations are restricted to 2,000 groups with 5 members resulting in a sample size of 10,000. It is conceivable that the

approximation of $\Lambda(t_l;\theta)$ improves with increasing sample size, which may in turn improve the stableness of the estimator. However, the question of whether this procedure performs well with larger sample sizes remains open.

5.4.3 The use of Stata's semi-parametric shared frailty estimator with left-truncated data in empirical work

Stata's `stcox` model has been frequently used in the empirical literature to estimate shared gamma frailty models, and sometimes the data are left-truncated. Gottard and Rampichini (2006) study the effects of poverty on time to childbirth among young women in Bolivia. In their data, individuals within a region are assumed to share their frailty term, and are only included in their sample if they have reached at least the age of 14 at the time of the survey in 1998. Hence, left-truncation points vary across individuals. They state that they use the `stcox, shared` command in their empirical analysis. Another example is provided by Studenski et al. (2011), who study the effect of gait speed on survival among elderly individuals. They use data from 9 different cohort studies and estimate shared gamma frailty models with Stata in a sensitivity analysis of their main results, with the frailty taken to be cohort-study-specific. The individual lifetime durations are left-truncated by the entry age into the study. Hemmelgarn et al. (2007) study multidisciplinary care for elderly patients with chronic kidney disease, including its effect on survival. They assume shared frailties for matched treated and untreated individuals, and estimate shared frailty models with Stata and/or SAS. Their data are subject to left-truncation. Matching on age ensures that both lifetime durations need to exceed a left-truncation point in order for the pair to be included in the sample.

5.5 Conclusion

This paper analyzes the implications of ignoring the effect of left-truncation of duration data on the distribution of unit-specific unobserved determinants in the sample, if multiple durations are observed per unit. In the presence of unobserved heterogeneity, it is vital to correctly account for the truncation that influences the composition of survivors in the sample, especially if the truncation thresholds vary across units.

Stata users estimating shared frailty models with the `streg` or `stcox` command need

to be aware that with left-truncated data, the estimators of the covariate effects, the duration dependence and the variance of the frailty distribution may be inconsistent. The magnitude of the bias depends on the level of truncation, as well as on the variance of the frailty distribution of the data generating process. The good news is that the parameter estimates for the covariate effects are typically biased towards zero. Therefore, in the worst case, effects have been underestimated by Stata.

5.A Appendix

First, note that the gamma and Inverse-Gaussian distributions are both special cases of the non-negative exponential family with density

$$f(v) = v^{\delta} e^{-\lambda v} m(v) \phi(\delta, \lambda)^{-1}. \tag{5.13}$$

A shared frailty model with a frailty distribution of this family has the following survival function (see Hougard, 2000):

$$\begin{aligned} S(t_1, t_2|x) &= \int_0^{\infty} v^{\delta} e^{-(\lambda + M(t_1,t_2))v} m(v)\, dv \frac{1}{\phi(\delta, \lambda)} \\ &= \frac{\phi(\delta, \lambda + M(t_1, t_2))}{\phi(\delta, \lambda)}, \end{aligned} \tag{5.14}$$

with $M(t_1, t_2) = \phi(x_1)\Lambda(t_1) + \phi(x_2)\Lambda(t_2)$. The second equality follows from the fact that (5.13) is equivalent to $\phi(\delta, \lambda) = \int_0^{\infty} v^{\delta} e^{-\lambda v} m(v)\, dv$ and therefore $\phi(\delta, \lambda + M(t_1, t_2)) = \int_0^{\infty} v^{\delta} e^{-(\lambda + M(t_1,t_2))v} m(v)\, dv$.

5.A.1 Gamma frailty

Let us assume a gamma distributed frailty with $E(v) = 1$ and $Var(v) = \sigma^2$. This implies the following restrictions on the density function in (5.13)

$$\delta = 1/\sigma^2 - 1, \quad \lambda = 1/\sigma^2, \quad m(v) = 1, \quad \phi(\delta, \lambda) = \lambda^{-(\delta+1)}\Gamma(\delta + 1), \tag{5.15}$$

where $\Gamma(\sigma^2)$ is the gamma function. Substituting the expression for $\phi(\delta, \lambda)$ into the right hand side of equation (5.14) leads to

$$\begin{aligned} S(t_1, t_2|x) &= \frac{(1/\sigma^2 + M(t_1, t_2))^{-1/\sigma^2}\Gamma(1/\sigma^2)}{1/\sigma^{2-1/\sigma^2}\Gamma(1/\sigma^2)} \\ &= (1 + \sigma^2 M(t_1, t_2))^{-1/\sigma^2}. \end{aligned} \tag{5.16}$$

Since $f(t_1, t_2|x) = \frac{\partial^2(1 - S(t_1,t_2|x))}{\partial t_1 \partial t_2}$ it follows

$$f(t_1, t_2|x) = \frac{\partial M(t_1, t_2)}{\partial t_1} \frac{\partial M(t_1, t_2)}{\partial t_2} (\sigma^2 + 1)(1 + \sigma^2 M(t_1, t_2))^{-(1/\sigma^2 + 2)}. \tag{5.17}$$

Finally, let us consider the likelihood contribution of a group i comprising two subjects with truncation points t_{01} and t_{02} and no censoring. Combining the results from equation (5.16) and (5.17) leads to

$$\begin{aligned} f(t_1, t_2|T_1 > t_{01}, T_2 > t_{02}, x) &= \frac{f(t_1, t_2|x)}{S(t_{01}, t_{02}|x)} \\ &= \phi(x_1)\lambda(t_1)\phi(x_2)\lambda(t_2)(\sigma^2+1)(1+\sigma^2 M(t_{01}, t_{02}))^{1/\sigma^2}(1+\sigma^2 M(t_1, t_2))^{-(1/\sigma^2+2)} \end{aligned}$$

which is equation (5.2) from section 2.

5.A.2 Likelihood function in the Stata Manual

The Stata Reference Manual (Stata, 2009, p.383) presents the following likelihood contribution for a group i of a shared frailty model with a gamma frailty in the case of no censoring

$$L = \phi(x_1)\lambda(t_1)\phi(x_2)\lambda(t_2)(\sigma^2+1)(1+\sigma^2(M(t_1, t_2) - M(t_{01}, t_{02})))^{-(1/\sigma^2+2)}.$$

Rearranging and choosing $\delta = 1/\sigma^2 - 1$ and $\lambda = 1/\sigma^2$ according to (5.15) yields

$$L = \phi(x_1)\lambda(t_1)\phi(x_2)\lambda(t_2)\frac{(\lambda + M(t_1, t_2) - M(t_{01}, t_{02}))^{-(\delta+3)}\Gamma(\delta+3)}{(\lambda)^{-(\delta+1)}\Gamma(\delta+1)}.$$

Since we know that $\phi(\delta+2, \lambda+x) = (\lambda+x)^{-(\delta+3)}\Gamma(\delta+3)$ from (5.15) and that $\phi(\delta+2, \lambda+x) = \int_0^\infty v^{\delta+2}e^{-(\lambda+x)v}m(v)\,dv$ from equation (5.14) it follows

$$L = \phi(x_1)\lambda(t_1)\phi(x_2)\lambda(t_2)\int_0^\infty v^2 e^{-(M(t_1,t_2)-M(t_{01},t_{02}))v}\frac{v^\delta e^{-\lambda v}m(v)}{\lambda^{-(\delta+1)}\Gamma(\delta+1)}\,dv$$

and once the restrictions (5.15) for the gamma distribution are imposed again

$$L = \int_0^\infty f(t_1, t_2|T_1 > t_{01}, T_2 > t_{02}, x, v)\,dG(v).$$

5.A.3 Inverse-Gaussian frailty

Let us assume Inverse-Gaussian distributed frailty terms. Like with the gamma frailty, this imposes restrictions on the density in (5.13)

$$\delta = -1/2, \quad m(v) = \psi^{1/2}\pi^{-1/2}e^{-\frac{\psi}{v}}v^{-1}, \quad \phi(-1/2,\lambda) = e^{-(4\psi\lambda)^{1/2}}.$$

Assuming $\psi = \lambda$ gives a mean frailty of 1 and choosing $\sigma^2 = 1/(2\lambda)$ yields $Var(v) = \sigma^2$. Substituting the expression for $\phi(\delta,\lambda)$ into the right hand side of equation (5.14) leads to

$$\begin{aligned} S(t_1,t_2|x) &= \frac{\exp(-(4(\frac{1}{2\sigma^2})(\frac{1}{2\sigma^2}+M(t_1,t_2)))^{1/2})}{\exp(-(4(\frac{1}{2\sigma^2})^2)^{\frac{1}{2}})} \\ &= \exp(1/\sigma^2 - 1/\sigma^2(1+2\sigma^2M(t_1,t_2))^{1/2}). \end{aligned} \tag{5.18}$$

Since $f(t_1,t_2|x) = \frac{\partial^2(1-S(t_1,t_2|x))}{\partial t_1 \partial t_2}$ it follows

$$f(t_1,t_2|x) = \frac{\partial M(t_1,t_2)}{\partial t_1}\frac{\partial M(t_1,t_2)}{\partial t_2}\frac{(1+\sigma^2(1+2\sigma^2M(t_1,t_2))^{-\frac{1}{2}})S(t_1,t_2|x)}{1+2\sigma^2M(t_1,t_2)}. \tag{5.19}$$

Finally, let us consider the likelihood contribution of a group i comprising two subjects with truncation points t_{01} and t_{02} and no censoring. Combining the results from equation (5.18) and (5.19) yields

$$\begin{aligned} &f(t_1,t_2|T_1>t_{01},T_2>t_{02},x) \\ &= \frac{f(t_1,t_2|x)}{S(t_{01},t_{02}|x)} \\ &= \phi(x_1)\lambda(t_1)\phi(x_2)\lambda(t_2) \\ &\quad \times \frac{(1+\sigma^2(1+2\sigma^2M(t_1,t_2))^{-\frac{1}{2}})\exp(1/\sigma^2 - 1/\sigma^2(1+2\sigma^2M(t_1,t_2))^{1/2})}{(1+2\sigma^2M(t_1,t_2))\exp(1/\sigma^2 - 1/\sigma^2(1+2\sigma^2M(t_{01},t_{02}))^{1/2})}. \end{aligned}$$

5.A.4 The command stregshared

Syntax

The command `stregshared` (see http://vandenberg.vwl.uni-mannheim.de/2999.0.html) is designed as an alternative to `streg` when fitting a shared gamma frailty model to left-truncated duration data. The size of the units over which the frailties are shared should not exceed two when using `stregshared`. The functional form of the baseline hazard can

be specified as piecewise constant, Weibull, exponential or Gompertz. The command has a similar syntax to `streg`:

`stregshared` *varlist* [`if`] [`in`], `shared`(*varname*) [`noconstant` `distribution`(*baseline*) `cuts`(*numlist*)]

Description

`stregshared` is implemented as a `v0` evaluator and uses Stata's modified Newton-Raphson maximization algorithm. The command fits the same shared frailty model as the `streg` command with the `shared()` option, with the only difference being the adjusted likelihood function described in Section 2. Like `streg`, it requires the data to be defined as duration data by `stset` and uses the same variables in the same format as input arguments as `streg`.

Options

`noconstant` suppresses the constant term. The default is to include a constant in the model. Note that *varlist* should not include a constant term, when the option `noconstant` is not used.

`distribution`(*baseline*) sets the baseline hazard function to be of the type *baseline*, where *baseline* can be specified as `weibull`, `exponential` or `gompertz`. If this option is not used, a Weibull model is estimated. Note that the piecewise constant model requires this option to be specified as `d(exponential)`.

`cuts`(*numlist*) specifies the cutoff points of a piecewise constant baseline hazard. When the options `noconstant` and `d(exponential)` are used, the option `cuts(`***numlist***`)` allows estimating a piecewise constant model. Here, *numlist* holds the list of cutoff points, where the numbers have to be in strictly ascending order. For example, if the baseline function should be piecewise constant on the intervals $[0, 5.5)$, $[5.5, 10)$ and $[10, \infty]$, use: `nocon d(exponential) cuts(5.5,10)`. The option `cuts()` cannot be used with `d(weibull)` or `d(gompertz)`.

`shared`(*varname*) specifies a variable defining the units within which the frailty is shared. The variable in *varname* is the same variable used in the option `shared` of `streg`. Recall that `stregshared` can only deal with a unit size of one or two spells. It is not a problem for the command if some (but not all) of the units only have one spell and others have two. However, it cannot be used with units holding more than two spells. The `shared()` option has to be specified.

Comparison to `streg`

Given that the `stregshared` command was designed as an alternative to `streg`, it is intended to work in a very similar way. Therefore, if one uses the original `streg` Stata command after `stset` to estimate a shared gamma-frailty model with a Weibull distribution

```
. stset duration, failure(fail == 1) enter(truncation)
. streg x1 x2 x3, shared(id) d(weibull) frailty(gamma) nohr
```

the same arguments can be used with the `stregshared` command in order to estimate the same model with the adjustment in the likelihood function from Section 2:

```
. stset duration, failure(fail == 1) enter(truncation)
. stregshared x1 x2 x3, shared(id) d(weibull)
```

Here, `id` is the variable that identifies the unit. The same variable is used in the option `shared()` in `streg`. Note that the option `nohr` which causes `streg` to display the estimated parameter values rather than of the hazard ratios, is not used in our command. `stregshared` will display the parameter values as well as the hazard ratios in the estimation results.

In this example, the data are left-truncated and thus the `enter(truncation)` option in `stset` is used, with `truncation` being the variable that holds the left-truncation points for each spell. If the `enter()` option is not used in `stset`, `stregshared` and `streg` will yield the same estimation results.

Saved Results

When an estimation is run with `stregshared`, the command shows the choice of baseline function, the starting values, the number of units and total observations used in the estimation and, finally, the estimation results. These results include the parameter estimates, standard deviations, values of the test statistics and the hazard ratios.

`stregshared` saves the following in `e()`:

Scalars :	
`est_base`	ancillary parameter (for Weibull or Gompertz function)
`est_theta`	frailty parameter
Matrices :	
`est_b`	coefficient vector
`est_matrix`	complete matrix of estimation results (estimates, std. err. and test statistics)

To display the matrix of estimation results after running `stregshared`, type:

```
matrix list e(est_matrix)
```

Bibliography

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part iii. In *Handbook of Econometrics, edition 1, volume 6*. Elsevier.

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*(3), 701–710.

Abbring, J. H. and G. J. van den Berg (2003b). The nonparametric identification of treatment effects in duration models. *Econometrica 71*(5), 1491–1517.

Abbring, J. H. and G. J. van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika 94*(1), 87–99.

Abbring, J. H., G. J. van den Berg, and J. C. van Ours (2005). The effect of unemployment insurance sanctions on the transition rate from unemployment to employment. *The Economic Journal 115*(505), 602–630.

Alter, K. (2002). Do International Courts Enhance Compliance with International Law? *Review of Asian and Pacific Studies 25*, 51–78.

Alter, K. (2006). Private litigants and the new international courts. *Comparative Political Studies 39*(1), 22–49.

Altonji, J. G., S. Cattan, and I. Ware (2010). Identifying sibling influence on teenage substance use. Technical report, National Bureau of Economic Research.

Ansart, S., C. Pelat, P. Boelle, F. Carrat, A. Flahault, and A. Valleron (2009). Mortality burden of the 1918-1919 influenza pandemic in europe. *Influenza and Other Respiratory Viruses 3*, 99–106.

Aral, S. and D. Walker (2012). Identifying influential and susceptible members of social networks. *Science 337*(6092), 337–341.

Arnold, C. (2013). A different kind of game: International cooperation in mercosur. Technical report, Nuffield College Working Papers in Politics.

Bagnoli, M. and B. Lipman (1989, October). Provision of Public Goods: Fully Implementing the Core through Private Contributions. *The Review of Economic Studies 56*(4), 583–601.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's who in networks. wanted: the key player. *Econometrica 74*(5), 1403–1417.

Bartholomew, D. J., F. Steele, I. Moustaki, and J. I. Galbraith (2008). *Analysis of Multivariate Social Science Data*. Boca Raton: Chapman & Hall/CRC.

Beck, N. and J. Katz (1996). Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models. *Political Analysis 6*(1), 36.

Beck, T., G. Clarke, A. Groff, P. Keefer, and P. Walsh (2001). New Tools in Comparative Political Economy: The Database of Political Institutions. *The World Bank Economic Review 15*(1), 165–176.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2010). Identification of social interactions. Economics Series 260, Institute for Advanced Studies.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review 85*(3), 365–90.

Bouzas, R., S. Gratius, H. Soltz, and S. Sberro (2008). Teoría y Práctica de las Instituciones y Procesos de Decisión. In S. Gratius (Ed.), *MERCOSUR y NAFTA. Instituciones y Mecanismos de Decisión en Procesos de Integración Asimetricos*, pp. 33–96. Madrid: Iberoamericana.

Bouzas, R. and H. Soltz (2001, June). Institutions and regional integration: The case of mercosur. In *Regional Integration in Latin America and the Caribbean: the Political Economy of Open Regionalism*, Volume , pp. 1–25. (London: Palgrave, forthcoming).

Bowling, A. (1987, January). Mortality after bereavement: A review of the literature on survival periods and factors affecting survival. *Social Science & Medicine 24*(2), 117–124.

Box-Steffensmeier, J. M. and B. Jones (1997). Time is of the essence: Event history models in political science. *American Journal of Political Science 41*(4), 1414–1461.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics 30*, 89–99.

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

Brock, W. A. and S. N. Durlauf (2001a). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Brock, W. A. and S. N. Durlauf (2001b). Interactions-based models. In J. Heckman and E. Leamer (Eds.), *Handbook of econometrics*, Volume 5, Chapter 54, pp. 3297 – 3380. Elsevier.

Büthe, T. and H. V. Milner (2008). The Politics of Foreign Direct Investment Into Developing Countries: Increasing FDI Through International Trade Agreements? *American Journal of Political Science 52*(4), 741–762.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Canudas-Romo, V. and A. Erlangsen (2008). Denmark: the lowest excess mortality during the influenza pandemic of 1918.

Carrubba, C. (2005). Courts and Compliance in International Regulatory Regimes. *The Journal of Politics 67*(3), 669–689.

Case, A. C. and L. F. Katz (1991, May). The company you keep: The effects of family and neighborhood on disadvantaged youths. Working Paper 3705, National Bureau of Economic Research.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. Heckman and B. Singer (Eds.), *Longitudinal analysis of labor market data*. Cambridge University Press.

Checkel, J. T. (2005). International Institutions and Socialization in Europe: Introduction and Framework. *International Organization 59*(3), 801–826.

Chen, M. (2010). Youth smoking onset and the sibling effect. Unpublished manuscript, University at Albany.

Clark, A. E. and Y. Loheac (2007). 'it wasn't me, it was them!' social influence in risky behavior by adolescents. *Journal of health economics 26*(4), 763–784.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*(1), 141–151.

Cleves, M. A., W. W. Gould, and R. G. Gutierrez (2008). *An introduction to survival analysis using Stata*. Stata Press.

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

Dai, X. (2006, October). The Conditional Nature of Democratic Compliance. *The Journal of Conflict Resolution 50*(5), 690–713.

Dai, X. (2007). *International Institutions and National Policies*. Cambridge: Cambridge University Press.

Do, P. and S. Ma (2010). Frailty model with spline estimated nonparametric hazard function. *Statistica Sinica 20*(2), 561.

Doblhammer, G. (2004). *The late life legacy of very early life*. Springer, Berlin.

Downs, G., D. Rocke, and P. Barsoom (1996). Is the good news about compliance good news about cooperation? *International Organization 50*(3), 379–406.

Drepper, B. and G. Effraimidis (2012). Nonparametric identification of dynamic treatment effects in competing risks models. Unpublished manuscript, University of Mannheim, Economics Department.

Eisenberg, D. (2004). *Peer effects for adolescent substance use: Do they really exist*. Ph. D. thesis, UC-Berkeley School of Public Health.

Elbers, C. and G. Ridder (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies 49*(3), 403–09.

Fearon, J. D. (1998). Bargaining, Enforcement, and International Cooperation. *International Organization 52*(2), 269–305.

Finnemore, M. and K. Sikkink (1998, October). International Norm Dynamics and Political Change. *International Organization 52*(4), 887–917.

Flourney, N. (1992). Discussion of assessment of dependence in the life times of twins. In J. Klein and P. Goel (Eds.), *Survival Analysis: State of the Art*, pp. 91–97. Kluwer Academic Publishers.

Gans, J. S., D. H. Hsu, and S. Stern (2008). The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. *Management Science 54*(5), 982–997.

Garrett, G. and B. Weingast (1993). Ideas, Interests, and Institutions: Constructing the European Community's Internal Market. In J. Goldstein and R. Keohane (Eds.), *Ideas and Foreign Policy*, pp. 173–206. Ithaca: Cornell University Press.

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

Gavrilov, L. A. and V. N. Nosov (1985). A new trend in human mortality decline: derectangularization of the survival curve. *Age 8*, 93.

Gilligan, M. J., L. Johns, and B. P. Rosendorff (2010, January). Strengthening International Courts and the Early Settlement of Disputes. *Journal of Conflict Resolution 54*(1), 5–38.

Goldstein, J., D. Rivers, and M. Tomz (2007). Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade. *International Organization 61*(4), 37–67.

Gómez-Mera, L. (2009). Domestic Constraints to Regional Cooperation: Explaining Trade Conflict in Mercosur. *Review of International Political Economy 47*(1), 1–32.

Goodliffe, J. and D. G. Hawkins (2006). Explaining Commitment: States and the Convention Against Torture. *The Journal of Politics 68*(2), 358–371.

Gottard, A. and C. Rampichini (2006). Shared frailty graphical survival models. In *International Conference on Statistical Latent Variables Models in the Health Sciences*.

Guo, G. (1992). Event-history analysis for left-truncated data. *Sociological methodology 23*, 217–243.

Gutierrez, R. G. (2002). Parametric frailty and shared frailty survival models. *The Stata Journal 2, Number 1*, 22–44.

Guzman, A. T. (2008). *How International Law Works*. Oxford: Oxford University Press.

Harris, J. E. and B. González López-Valcárcel (2008). Asymmetric peer effects in the analysis of cigarette smoking among young people in the united states, 1992–1999. *Journal of Health Economics 27*(2), 249–264.

Hathaway, O. (2003, May). The Cost of Commitment. *Stanford Law Review 55*(5), 1821–1862.

Hathaway, O. (2005, April). Between Power and Principle: An Integrated Theory of International Law. *The University of Chicago Law Review 72*(2), 469–536.

Hathaway, O. A. (2007). Why Do Countries Commit to Human Rights Treaties? *Journal of Conflict Resolution 51*(4), 588–621.

Heckathorn, D. (1996, April). The Dynamics and Dilemmas of Collective Action. *American Sociological Review 61*(2), 250–277.

Heckman, J. J. and B. E. Honoré (1989). The identifiability of the competing risks model. *Review of Economic Studies 77*, 1138–1163.

Heckman, J. J. and C. R. Taber (1994). Econometric mixture models and more general models for unobservables in duration analysis. *Statistical Methods in Medical Research 3*(3), 279–299.

Hemmelgarn, B. R., B. J. Manns, J. Zhang, M. Tonelli, S. Klarenbach, M. Walsh, B. F. Culleton, et al. (2007). Association between multidisciplinary care and survival for elderly patients with chronic kidney disease. *Journal of the American Society of Nephrology 18*(3), 993–999.

Hirsch, K. and A. Wienke (2012). Software for semiparametric shared gamma and log-normal frailty models: An overview. *Computer methods and programs in biomedicine 107*(3), 582–597.

Holm, N. V. (1983). *The use of twin studies to investigate causes of diseases with complex etiology with a focus on cancer*. Ph. D. thesis, Odense University.

Honoré, B. and A. de Paula (2013). Interdependent durations in joint retirement. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Honoré, B. E. and A. de Paula (2010). Interdependent durations. *Review of Economic Studies 77*, 1138–1163.

Hougaard, P., B. Harvald, and N. V. Holm (1992a). Assessment of dependence in the life times of twins. In J. Klein and P. Goel (Eds.), *Survival Analysis: State of the Art*, pp. 77–90. Kluwer Academic Publishers.

Hougaard, P., B. Harvald, and N. V. Holm (1992b, March). Measuring the similarities between the lifetimes of adult danish twins born between 1881-1930. *Journal of the American Statistical Association 87*(417), 17–24.

Hougard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

Iyengar, R., S. Han, and S. Gupta (2009). Do friends influence purchases in a social network? *Harvard Business School Marketing Unit Working Paper* (09-123).

Jacobson, H. K. (2000). International Institutions and System Transformation. *Annual Review of Political Science 3*, 149–166.

Jensen, H., R. B., P. Aaby, and P. Andersen (2004). *Shared frailty model for left-truncated multivariate survival data*. Department of Biostatistics, University of Copenhagen.

Johnson, V. and J. Albert (1999). *Ordinal Data Modelling*. New York: Springer.

Joliffe, I. T. (2002). *Principal Component Analysis* (2nd Edition ed.). New-York: Springer.

Kalbfleisch, J. D. and R. L. Prentice (2011). *The statistical analysis of failure time data*, Volume 360. John Wiley & Sons.

Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American journal of Sociology*, 427–436.

Kawaguchi, D. (2004). Peer effects on substance use among american teenagers. *Journal of Population Economics 17*(2), 351–367.

Keohane, R. (1984). *After hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.

Keohane, R., A. Moravcsik, and A.-M. Slaughter (2000). Legalized Dispute Resolution: Interstate and Transnational. *International Organization 54*(3), 457–488.

Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics 48*, 795–806.

Kono, D. Y. (2006, August). Optimal Obfuscation: Democracy and Trade Policy Transparency. *American Political Science Review 100*(3), 369–384.

Kono, D. Y. (2007). Making Anarchy Work: International Legal Institutions and Trade Cooperation. *The Journal of Politics 69*(3), 746–759.

Koremenos, B., C. Lipson, and D. Snidal (2001). The Rational Design of International Institutions. *International Organization 55*(4), 761–799.

Kortram, R. A., A. C. M. van Rooij, A. J. Lenstra, and G. Ridder (1995a). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica 49*(3), 269–281.

Kortram, R. A., A. C. M. van Rooij, A. J. Lenstra, and G. Ridder (1995b). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica 49*(3), 269–281.

Kragt, A., J. Orbell, and R. Dawes (1983, March). The Minimal Contributing Set as a Solution to Public Goods Problems. *American Political Science Review 77*(1), 112–122.

Krapohl, S., J. Dinkel, and B. Faude (2009, June). Judicial Integration in the Americas. *Working Paper University of Bamberg*, 1–30.

Krasner, S. D. (1991, April). Global Communications and National Power: Life on the Pareto Frontier. *World Politics 43*(3), 336–366.

Krauth, B. V. (2005). Peer effects and selection effects on smoking among canadian youth. *Canadian Journal of Economics 38*(3), 735–757.

Lalive, R., J. C. van Ours, and J. Zweimüller (2005). The effect of benefit sanctions on the duration of unemployment. *Journal of the European Economic Association 3*(6), 1386–1417.

Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge: Cambridge University Press.

Lancaster, T. (1992). *The econometric analysis of transition data*. Number 17. Cambridge university press.

Lavopa, J. H. (2003). *Derecho Constitucional Y Procesos De Integración. La Integración en Los Estados Federales Del MERCOSUR: El Caso Argentino Y Brasileño.* Anuario de Derecho Constitucional Latinoamericano. Mexico City: Fundación Konrad Adenauer.

Lenz, T. (2012). Spurred Emulation: the EU and Regional Integration in MERCOSUR and SADC. *West European Politics 35*(1), 155–173.

Lichtenstein, P., M. Gatz, and S. Berg (1998, 5). A twin study of mortality after spousal bereavement. *Psychological Medicine 28*, 635–643.

Lindeboom, M., F. Portrait, and G. J. van den Berg (2002). An econometric analysis of the mental-health effects of major events in the life of older individuals. *Health Economics 11*(6), 505–520.

Long, A., T. Nordstrom, and K. Baek (2007). Allying for Peace: Treaty Obligations and Conflict between Allies. *The Journal of Politics 69*(4), 1103–1117.

Lykken, D. T. (1978). The diagnosis of zygosity in twins. *Behavior Genetics 8*, 437–473.

Malamud, A. (2005). Presidential Diplomacy and the Institutional Underpinnings of Mercosur: An Empirical Examination. *Latin American Research Review 40*(1), 138–164.

Manor, O. and Z. Eisenbach (2003). Mortality after spousal loss: are there socio-demographic differences? *Social Science & Medicine 56*(2), 405 – 413.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Manski, C. F. (2000). Economic analysis of social interactions. *The Journal of Economic Perspectives 14*(3), 115–136.

Martin, L. (2000). *Democratic Commitments: Legislatures and International Cooperation.* Princeton: Princeton University Press.

Martin, L. (2005, September). The President and International Commitments: Treaties as Signaling Devices. *Presidential Studies Quarterly 35*(3), 440–465.

Martin, L. (2008). Xinyuan Dai, International Institutions and National Policies. *The Review of International Organizations*.

McVicar, D. and A. Polanski (2013). Peer effects in uk adolescent substance use: Never mind the classmates? *Oxford Bulletin of Economics and Statistics*.

Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions. In *Social dynamics*, pp. 45–82. Cambridge: MIT Press.

Monstad, K., C. Propper, and K. Salvanes (2011). Is teenage motherhood contagious? evidence from a natural experiment.

Moriarty, J., D. McVicar, and K. Higgins (2012). Peer effects in adolescent cannabis use: It's the friends, stupid.

National Longitudinal Surveys Handbook (2005). *The NLSY79* (Chapter 3, 34 - 52 ed.). National Longitudinal Surveys Handbook.

Nielsen, G. G., R. D. Gill, P. K. Andersen, and T. I. Sørensen (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 25–43.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory Of Groups*. Cambridge: Harvard University Press.

Ostrom, E. (1990). *Governing the Commons*. New York: Cambridge University Press.

Ostrom, E. (2003). How types of goods and property rights jointly affect collective action. *Journal of Theoretical Politics 15*(3), 239–270.

Oye, K. A. (1985, October). Explaining Cooperation under Anarchy: Hypotheses and Strategies. *World Politics 38*(1), 1–24.

Parner, E. (1997). *Inference in semiparametric frailty models*. Ph. D. thesis, Aarhus University, Department of Public Health-Institute of Biostatistics.

Pena, C. and R. Rozemberg (2005). Mercosur: A Different Approach to Institutional Development. *FOCAL Policy Paper 5*(6).

Peto, R. (1972). Contribution to the discussion of paper by d. r. cox. *Journal of the Royal Statistical Society. Series B 34*, 187–220.

Powell, E. and J. Staton (2009). Domestic Judicial Institutions and Human Rights Treaty Violation. *International Studies Quarterly 53*(1), 149–174.

Pudney, S. (2002). *The Road to Ruin?: Sequences of initiation into drug use and offending by young people in Britain*. Home Office London.

Ridder, G. (1984). *The distribution of single-spell duration data*. Springer.

Ridder, G. and I. Tunali (1999, October). Stratified partial likelihood estimation. *Journal of Econometrics 92*(2), 193–232.

Ridder, G. and T. M. Woutersen (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica 71*(5), 1579–1589.

Risse, T., S. C. Ropp, and K. Sikkink (1999). The power of human rights. *International Norms and Domestic Change, Cambridge*.

Rondeau, V. and J. R. Gonzalez (2005). Frailtypack: A computer program for the analysis of correlated failure time data using penalized likelihood estimation. *Computer methods and programs in biomedicine 80*(2), 154–164.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Sampson, R. J., J. D. Morenoff, and T. Gannon-Rowley (2002). Assessing 'neighborhood effects': Social processes and new directions in research. *Annual review of sociology*, 443–478.

Sanders, C. M. (1980). A comparison of adult bereavement in the death of a spouse, child, and parent. *Journal of Death and Dying 10*(4), 303–322.

Sanders, C. M. (1999). *Grief: The Mourning After: Dealing with Adult Bereavement, 2nd Edition*. Wiley & Sons, Chichester.

Scharpf, F. W. (1997). *Games Real Actors Play. Actor-Centered Institutionalism in Policy Research*. Boulder: Westview Press.

Schram, A., T. Offerman, and J. Sonnemans (2008). *Handbook of Experimental Economics Results*, Volume 1. Elsevier.

Segal, N. L. and S. A. Blozis (2002). Psychobiological and evolutionary perspectives on coping and health characteristics following loss: a twin study. *Twin Research 5*, 175–87.

Segal, N. L. and T. J. Bouchard (1993). Grief intensity following the loss of a twin and other relatives: test of kinship genetic hypotheses. *Human Biology 65*, 87–105.

Segal, N. L. and S. L. Ream (1998). Decrease in grief intensity for deceased twin and non-twin relatives: an evolutionary perspective. *Personality and Individual Differences 25*(2), 317 – 325.

Segal, N. L., L. J. Sussman, W. D. Marelich, J. Mearns, and S. A. Blozis (2002, 6). Monozygotic and dizygotic twins' retrospective and current bereavement-related behaviors: An evolutionary perspective. *Twin Research 5*, 188–195.

Segal, N. L., S. M. Wilson, T. J. Bouchard, and D. G. Gitlin (1995). Comparative grief experiences of bereaved twins and other bereaved relatives. *Personality and Individual Differences 18*(4), 511 – 524.

Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature 138*, 32.

Selye, H. (1955). Stress and disease. *Science 122*(3171), 625–631.

Simmons, B. A. (2000, December). International Law and State Behavior: Commitment and Compliance in International Monetary Affairs. *American Political Science Review 94*(4), 819–835.

Simmons, B. A. and A. Danner (2010). Credible Commitments and the International Criminal Court. *International Organization 64*(1), 225–256.

Skytthe, A., K. Kyvik, N. Holm, J. Vaupel, and K. Christensen (2002). The danish twin registry: 127 birth cohorts of twins. *Twin Research 5*, 352–357.

Smith, Kline, F. (1969). *Emphasis: the Smith Kline & French magazine*. Philadelphia, Pa: Smith Kline & French Laboratories.

Snidal, D. (1985, December). Coordination versus Prisoners' Dilemma: Implications for International Cooperation and Regimes. *American Political Science Review 79*(4), 923–942.

Stata (2009). Stata survival analysis and epidemiological tables, reference manual release 11. Technical report, Stata Press.

Stein, A. A. (1980). The Politics of Linkage. *World Politics 33*(1), 62–81.

Stein, A. A. (1982, April). Coordination and Collaboration: Regimes in an Anarchic World. *International Organization 36*(2), 299–324.

Stein, A. A. (1983). Coordination and Collaboration: Regimes in an Anarchic World. pp. 115–141. Ithaca: Cornell University Press.

Stroebe, M. S. and W. Stroebe (1987). *Bereavement and Health: The Psychological and Physical Consequences of Partner Loss*. Cambridge University Press.

Stroebe, M. S., W. Stroebe, and R. O. Hansson (1993). The mortality of bereavement: A review. In *Handbook of bereavement: Theory, research and intervention*, pp. 175 – 195. Cambridge University.

Studenski, S., S. Perera, K. Patel, C. Rosano, K. Faulkner, M. Inzitari, J. Brach, J. Chandler, P. Cawthon, E. B. Connor, M. Nevitt, M. Visser, S. Kritchevsky, S. Badinelli, T. Harris, A. B. Newman, J. Cauley, L. Ferrucci, and J. Guralnik (2011). Gait speed and survival in older adults. *Journal of the American Medical Association 305*(1), 50–58.

Svarer, M. and M. Verner (2008). Do children stabilize relationships in denmark? *Journal of Population Economics 21*(2), 395–417.

Therneau, T. and P. Grambsch (2000). *Modeling survival data: extending the Cox model*. Springer, New York.

Tollison, R. D. and T. D. Willett (1979). An Economic Theory of Mutually Advantageous Issue Linkages in International Negotiations. *International Organization 33*(04), 425.

Tomassini, C., F. C. Billari, A. Rosina, A. Skytthe, and K. Christensen (2001). Born together-die together. live together-die together. the role of the partner and of the co-twin on longevity at very old ages. *Genus* (3-4), 63–82.

Tomassini, C., A. Rosina, F. C. Billari, A. Skytthe, and K. Christensen (2002). The effect of losing the twin and losing the partner on mortality. *Twin Research 5*, 210–217.

Van den Berg, G. J. (2001). Chapter 55 duration models: specification, identification and multiple durations. In J. Heckman and E. Leamer (Eds.), *Handbook of econometrics*, Volume 5, pp. 3381 – 3460. Elsevier.

Van den Berg, G. J., M. Lindeboom, and F. Portrait (2011). Conjugal bereavement effects on health and mortality at advanced ages. *Journal of Health Economics 30*(4), 774 – 794.

Van den Berg, G. J., B. Van der Klaauw, and J. C. Van Ours (2004). Punitive sanctions and the transition rate from welfare to work. *Journal of Labor Economics 22*(1), 211–241.

Van Ours, J. C. (2003). Is cannabis a stepping-stone for cocaine? *Journal of Health Economics 22*(4), 539 – 554.

Van Ours, J. C. and J. Williams (2009). Why parents worry: Initiation into cannabis use by youth and their educational attainment. *Journal of Health Economics 28*(1), 132 – 142.

Ventura, D. and A. Perotti (2004). *El Proceso Legislativo Del Mercosur*. Montevideo: Comisión Parlamentaria Conjunta del Mercosur.

Von Stein, J. (2005, November). Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance. *American Political Science Review 99*(4), 611–622.

Whitehead, L. and A. Barahona de Brito (2005). Las Cumbres Mundiales Y Sus Versiones Latinoamericanas: Una Monta De Un Grano De Arena? *Arica Latina Hoy* (40), 15–27.

Widder, D. V. (1946). *The Laplace Transform*. Princeton University Press, Princeton.

Wienke, A., K. Christensen, A. Skytthe, and A. I. Yashin (2002). Genetic analysis of cause of death in a mixture model of bivariate lifetime data. *Statistical Modelling 2*, 89–102.

Wienke, A., N. Holm, A. Skytthe, and A. Yashin (2001, August). The heritability of mortality due to heart diseases: a correlated frailty model applied to danish twins. *Twin Research 4*, 266–74.

Williamson, O. (1975). *Markets and Hierarchies: Analysis and Antritrust Implications*. New York: Free Press.

Woodward, J. (1988). The bereaved twin. *Acta Geneticae Medicae et Gemellologiae 37*, 173–80.

Yarbrough, B. and R. Yarbrough (1997). Dispute Settlement in International Trade: Regionalism and Procedural Coordination. In E. D. Mansfield and H. V. Milner (Eds.), *The Political Economy of Regionalism*, pp. 134–163. New York: Columbia University Press.

Yashin, A. I. and I. A. Iachine (1995a). Genetic analysis of durations: Correlated frailty model applied to survival of danish twins. *Genetic Epidemiology 12*, 529–538.

Yashin, A. I. and I. A. Iachine (1995b). How long can humans live? lower bound for biological limit of human longevity calculated from danish tiwn data using correlated frailty model. *Mechanisms of Ageing and Development 80*, 147–169.

Young, H. P. (1991). *Negotiation analysis*. University of Michigan Press.

Curriculum Vitae

2013 onwards	Assistant professor, Tilburg University, Netherlands
2008 - 2013	PhD studies in Economics at the Center for Doctoral Studies in Economics, University of Mannheim, Germany Dissertation: The Multivariate Mixed Proportional Hazard Model: Applications and Extensions, Supervisor: Prof. Gerard van den Berg, Ph.D.
Autumn 2007	Research assistant at the Economic and Social Research Institute, Dublin, Ireland
Winter 2005/06	Visiting student in the Econometrics and Operations Research Master Program, Maastricht University, Netherlands
2002 - 2008	Diploma studies (M.A. equivalent) in economics at the Rheinische Friedrich-Wilhelms-Universität Bonn, Germany