DISCUSSION PAPER SERIES

IZA DP No. 810

A Simple Procedure for the Evaluation of Treatment Effects on Duration Variables

Jaap H. Abbring
Gerard J. van den Berg

June 2003

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

A Simple Procedure for the Evaluation of Treatment Effects on Duration Variables

Jaap H. Abbring
Free University of Amsterdam and IZA Bonn

Gerard J. van den Berg
Free University of Amsterdam, Tinbergen Institute, IFAU-Uppsala, INSEE-CREST, CEPR and IZA Bonn

Discussion Paper No. 810
June 2003

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

IZA Discussion Paper No. 810
June 2003

ABSTRACT

A Simple Procedure for the Evaluation of Treatment Effects on Duration Variables*

Often, a treatment and the outcome of interest are characterized by the moment they occur, and these moments are realizations of stochastic processes with dependent unobserved determinants. We develop a simple and intuitive method for inference on the treatment effect. The method can be implemented as a graphical procedure or as a straightforward parameter test in an auxiliary univariate single-spell duration model. The method exploits information on the timing of the treatment relative to the outcome that is discarded in binary treatment analyses.

JEL Classification: C14, C31, C41

Keywords: duration analysis, hazard rate, selectivity bias, treatment effect, unobserved heterogeneity

Corresponding author:

Jaap H. Abbring
Department of Economics
Free University of Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
Email: jabbring@econ.vu.nl

* Thanks to Arthur Lewbel and Bas van der Klaauw for useful suggestions. Also thanks to seminar participants at IFAU Uppsala, SSE Stockholm, Mannheim, Bristol, INSEE-CREST, and Amsterdam, and participants at ESPE, CESG and ESHE conferences. Jaap Abbring worked on this paper while visiting The University of Chicago, University College London and the Institute for Fiscal Studies. Abbring's research is supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

1 Introduction

In virtually every real-life treatment situation, both the treatment and the outcome of interest are realized at specific points in time. Examples in economics include the effect of training programs or punitive benefits reductions on unemployment durations, the effect of the hiring of replacement workers on strike durations, and the effect of promotions on tenure. Typically, the empirical analysis of the treatment effect is hampered by selection problems: individuals who obtain a treatment may have systematically different outcomes than those who do not. This creates non-trivial problems for inference. An extensive literature on the evaluation of treatment effects exists, but in general this literature does not address the specific timing of events in conjunction with selection problems (see for instance Heckman, LaLonde and Smith, 1999, for an overview of this literature). This paper is concerned with the evaluation of treatment effects in such a dynamic context. Consider a subject in a certain state. After a certain stochastic amount of time, the subject leaves this state. The subject may receive a treatment at some stochastic moment before it leaves the state. We are interested in the effect of the treatment on the duration in the state, or, equivalently, on the exit rate out of the state.

We develop a user-friendly method for obtaining evidence on the presence of a treatment effect from single-spell data that allows for possible selection effects. The method focuses on the sign of the treatment effect and not on its magnitude. We provide two complementary implementations of our method. First, it may be implemented as a graphical procedure. At first sight it may seem difficult to obtain a simple graphical representation of the sign of a treatment effect. A natural starting point would be to consider all individuals who have not yet received a treatment at a given point of time and who have not left the state of interest, and to compare those who do get a treatment to those who do not. However, a difference in the level of the exit rate out of this state may be due to the fact that one examines selected subsets of individuals. Alternatively, by analogy to difference-in-difference estimators of treatment effects, one may condition on the moment of treatment and compare what happens before and after this moment. However, to observe the moment of treatment one has to condition on not having left the state of interest before that moment, so one does not observe the exit rate out of the state of interest before the moment of treatment. In this paper we condition instead on the moment of exit and to examine the rate at which a treatment is given before that. We then focus on properties of this rate as a function of time, for different moments of exit, and we demonstrate that these rates are informative on the treatment effect. Basically, one has to examine whether the treatment rate increases more for individuals with a short stay in the state of interest than for individuals with a long stay in that state.

The second implementation of our method amounts to the estimation of an auxiliary parametric single-spell duration model for the conditional rates at which treatments are given. The estimates allow for a simple parameter test. Both implementations involve only straightforward manipulation of raw data. In both cases, it is the interaction between the moment of exit and the moment of treatment in the conditional treatment rate that allows one to distinguish between a causal treatment effect and selectivity.

Our point of departure is a nonparametric bivariate hazard rate model for the treatment and outcome hazard rates. Both hazard rates are taken to be multiplicative in duration dependence and unobserved heterogeneity terms, and a causal effect of the realized treatment works on the hazard rate for the outcome of interest from the moment the treatment is realized onwards. We do not exploit variation in observed covariates. If such covariates are available, the method can be applied to subsamples stratified with respect to the covariates. As such, our analysis implicitly allows for maximum interaction between duration dependence, unobserved heterogeneity and treatment effects on the one hand and the observed covariates on the other hand. This is a major advantage over the usual approach to identification of single-spell duration models, which rests heavily on covariate variation and assumptions with respect to the covariate effects (see Van den Berg, 2001, for an overview). Alternatively, one may discard covariates, in which case they can be absorbed by the unobserved heterogeneity term.

As a result, our method of inference does not require exclusion restrictions on observed covariates, so the data need not contain an explanatory variable that affects the treatment assignment but does not affect the outcome of interest other than by way of the treatment. Also, our method does not need conditional independence assumptions stating that the data capture all systematic determinants of the process of treatment assignment so that the remaining observed variation in the treatment assignment is independent of the determinants of the outcome of interest. Both exclusion restrictions and conditional independence assumptions are often difficult to justify.[1] Standard methods of treatment evaluation often

[1]If a variable is observed by the analyst then it is often also observable to the individuals under consideration. If the variable affects the probability of treatment, and the individual knows that he may be subject to treatment, then he takes his value of the variable into account to determine his optimal strategy, and this strategy affects the rate at which the individual leaves the state of interest. Indeed, if the individual knows that the variable is an important determinant of the treatment assignment process then he may have a strong incentive to in-

rely on exclusion restrictions, parametric functional form assumptions on the joint distribution of the "error terms" in the model, or conditional independence assumptions, to identify the treatment effect. In this sense, our method compares favorably to those methods.

At stated above, it is the interaction between the moment of exit and the moment of treatment in the conditional treatment rate that allows one to distinguish between a causal treatment effect and selectivity. Difference-in-differences methods of treatment evaluation in regression-type models[2] also use interactions in the data to make inference on treatment effects, and these methods also do not require observed covariates (the main difference with our approach is that we compare different individuals over time when quantifying the interaction effect.) Our results therefore illustrate the usefulness of the information in the timing of events to assess the treatment effect. We return to this in Subsection 4.1 below.

The paper is organized as follows. In Section 2 we discuss the model framework. Section 3 contains the analysis underlying our method. Section 4 discusses the implementation of the method. Section 5 briefly examines an alternative graphical procedure. Section 6 concludes.

2 The model framework

For the sake of convenience, we use the term "individual" in general to denote a subject in a state of interest. We normalize the point of time at which the individual enters the state to zero. The durations T_m and T_p measure the duration until the event of interest and the duration until treatment, respectively. The population that we consider concerns the inflow into the state, and the unconditional probability distributions that are defined below are distributions in the inflow into the state. Whether this is the inflow at a fixed point of calendar time or the total inflow over time (or the inflow at another range of inflow dates) depends on the application at hand.

The two durations are random variables. We use t_m and t_p to denote their realizations. We assume that, for a given individual in the population, the duration variables are absolutely-continuous random variables. We assume that the

quire the actual value of the variable. All of this is inconsistent with exclusion restrictions. Conditional independence may be difficult to justify if the treatment assignment is carried out by case workers who use discretionary power, taking individual characteristics of the subject into account that are unobserved to the analyst.

[2]See Angrist and Krueger (1999), Blundell and MaCurdy (1999), and Heckman, LaLonde and Smith (1999), for overviews.

effect of individual differences on the joint distribution of T_m, T_p can be captured by explanatory variables V that are unobserved to the analyst. As discussed in the introduction, we do not explicitly include observed covariates X in our analysis. It is implicitly understood that all our results are valid conditional on such covariates and therefore extend to a model that allows for arbitrary effects of X on the joint duration distribution and for dependence of V on X. Note that this implies that we do not exploit covariates for inference on the treatment effect.

Of course, the joint distribution of $T_m, T_p|V$ can be expressed in terms of the distributions of $T_p|V$ and $T_m|T_p = t_p, V$. The latter distributions are in turn characterized by their hazard rates $\theta_p(t|V)$ and $\theta_m(t|t_p, V)$, respectively.[3]

As noted in the introduction, we are interested in the causal effect of treatment on the exit out of the current state (see Abbring and Van den Berg, 2003, for a presentation in terms of potential-outcome variables). The treatment and the exit are characterized by the *moments* at which they occur, and we are interested in the effect of the realization of T_p on the distribution of T_m. To proceed, we assume that, conditional on V, the relation between T_m and T_p is characterized as follows: the realization of T_p affects the shape of the hazard of T_m from t_p onwards, in a deterministic way. In Abbring and Van den Berg (2003), this fundamental assumption is discussed in great detail. It implies that the causal effect is captured by the effect of t_p on $\theta_m(t|t_p, V)$ for $t > t_p$. Note that it is ruled out that t_p affects $\theta_m(t|t_p, V)$ on $t \in [0, t_p]$. In a behavioral model, a natural interpretation of this "‘no-feedback” assumption is that it excludes anticipation effects.[4]

[3] For a nonnegative random (duration) variable T, the hazard rate is defined as $\theta(t) = \lim_{dt \downarrow 0} \Pr(T \in [t, t + dt)|T \geq t)/dt$. Somewhat loosely, this is the rate at which the spell is completed at t given that it has not been completed before, as a function of t. It provides a full characterization of the distribution of T (see Lancaster, 1990, and Van den Berg, 2001). Consider the distribution of a duration variable conditional on some other variables. It is customary to use a vertical “conditioning line” within the argument of a hazard rate in order to distinguish between (on the left-hand side) the value of the duration variable at which the hazard rate is evaluated, and (on the right-hand side) the variables that are conditioned upon.

[4] In reality, there is often no information available on the degree to which an actual treatment is anticipated. Even if some anticipation cannot be ruled out, there is virtually never any information on the moment at which the individual receives information on the moment of treatment unless the moment of treatment is fully predictable at the individual level. The fact that a realization of the event of interest could be due to the anticipation of a future treatment has haunted the empirical literature on treatment effects. Many standard treatment evaluation studies suffer from a potential bias due to anticipatory effects. This includes studies using “difference-in-differences” methods where one “difference” concerns a comparison between pre- and post-treatment circumstances (see Heckman, LaLonde and Smith, 1999, for an overview).

Now suppose that the determinants of the stochastic process of treatment assignment affect the individual’s exit rate out of the state of interest before the actual realization of the

Let $V := (V_m, V_p)'$ be a (2×1)-vector of unobserved covariates, with a distribution G (in the inflow) such that $\Pr(0 < V_m < \infty, 0 < V_p < \infty) = 1$. Let $T_p \perp\!\!\!\perp V_m | V_p$, implying that $\theta_p(t|V) = \theta_p(t|V_p)$. Furthermore, let $T_m \perp\!\!\!\perp V_p | T_p, V_m$, so that $\theta_m(t|t_p, V) = \theta_m(t|t_p, V_m)$. Somewhat loosely, one may say that V_p (V_m) captures the unobserved determinants of T_p (T_m). Now let us turn to the specifications of the hazard rates $\theta_m(t|t_p, V_m)$ and $\theta_p(t|V_p)$. We adopt the following model:

$$
\begin{aligned}
&\theta_p(t|V_p) = \lambda_p(t) V_p \\
&\theta_m(t|t_p, V_m) = \lambda_m(t) \delta^{\mathrm{I}(t>t_p)} V_m,
\end{aligned}
\tag{1}
$$

where I(.) denotes the indicator function, which is 1 if its argument is true and 0 otherwise.

The functions $\lambda_m : [0, \infty) \to (0, \infty)$ and $\lambda_p : [0, \infty) \to (0, \infty)$ are called "baseline hazards". A hazard rate is said to be duration dependent if its value changes over t. Positive (negative) duration dependence means that $\lambda_i(t)$ increases (decreases). For expositional convenience only, we assume that λ_m and λ_p are differentiable on $(0, \infty)$. Note that this does not exclude that $\lim_{t \downarrow 0} \lambda_i(t) = \infty$ for $i = m$ and/or $i = p$, which implies that the popular Weibull specification $\lambda_i(t) = \alpha_i t^{\alpha_i - 1}$ with $\alpha_i > 0$ for the baseline hazards is included.[5] We also assume that $\Lambda_i(t) := \int_0^t \lambda_i(\tau) d\tau < \infty$ $(i = m, p)$ for every $t \geq 0$. Finally, we assume that $\lim_{t \to \infty} \Lambda_m(t) = \infty$. As we have also excluded a mass point at 0 in the distribution of V_m, this implies that the distribution of T_m cannot be defective. The model does allow for a defective distribution of T_p, through the possibility of $\lim_{t \to \infty} \Lambda_p(t) < \infty$. In the latter case each individual has a positive probability of never receiving a treatment even if they never leave the state of interest. Because of the assumptions on λ_p and V_p, the distributions of T_p and $T_p|V_p$ are not degenerate. This implies that at the individual level there is variation in the moment of treatment.

treatment. Then the treatment program is said to have an *ex ante* effect on exit out of the state of interest. Such an effect is to be expected in well-established programs. The ex ante effect should not be confused with anticipation of the *realization* of the process of treatment assignment, because in the latter case the individual knows the stochastic outcome rather than the determinants of the process. The ex ante effect can be contrasted to the *ex post* effect of treatment, which is the effect of a realized treatment on the individual exit rate – this is of course the effect we focus on in this paper.

[5] In that case, the value of λ_i *at* 0 needs to be an arbitrary positive number.

The term $\delta^{\mathrm{I}(t>t_p)}$ captures the treatment effect. Clearly, treatment is ineffective if and only if $\delta = 1$. Now suppose that $\delta > 1$. If T_p is realized then the level of the individual exit rate out of the current state increases by a fixed factor. This stochastically reduces the remaining duration in that state, in comparison to the case where the treatment is given at a later point of time.

The model does not impose parametric functional form assumptions on the baseline hazards or the probability distribution of the unobserved heterogeneity terms. As we allow for full interaction with observed covariates X, we do not impose that there are X that do affect T_p but do not affect T_m other than by way of t_p, so we do not impose an exclusion restriction.

Many empirical studies have estimated models that are closely similar to our model framework. All of these include effects of observed covariates as additional multiplicative terms in $\theta_m(t|t_p, V_m)$ and $\theta_p(t|V_p)$. For example, Card and Sullivan (1988), Gritz (1993), Bonnal, Fougère and Sérandon (1997), Abbring, Van den Berg and Van Ours (1997), and Van den Berg, Van der Klaauw and Van Ours (2004) study the effect of a treatment of unemployed workers on the transition rate from unemployment to work. Lillard (1993) estimates a model for the joint durations of marriage and time until conception of a child, and his model allows the rate at which the marriage dissolves to shift to another level at moments of child birth. Our model then describes a part of a two-dimensional stochastic process in which the occurrence of an event in one dimension affects the hazard rate in the other dimension. Lillard and Panis (1996) estimate a model on the joint durations of marriage, non-marriage, and life, and their model allows the death rate to shift to another level at moments of marriage formation and dissolution. Abbring and Van den Berg (2003) prove full identification of models with covariates and with treatment effects that depend on model variables. They also provide ample discussion of the model as an econometric model of treatment effects.

If we abstract from duration dependence and unobserved heterogeneity terms then the model is essentially equivalent to the bivariate exponential distribution developed by Freund (1961). This model is usually motivated by the example of a machine with two components, where one component may fail at a higher rate if the other component has already failed. The Freund (1961) model is symmetric to the extent that it allows each component to be affected by the failure of the other. However, if the observational plan is such that observation stops at the moment at which one specific component fails then this model is observationally equivalent to our model with constant baseline hazards, absence of unobserved heterogeneity, and constant δ. Similarly, our model can be extended to a symmetric setup by specifying the distribution of $[T_p|T_m = t_m, T_p \geq t_m]$. Note that the assumption that failure of one component induces a *higher* failure rate for the other means that $\delta > 1$.

We now discuss the general problem of inference on the treatment effect in the context of our model. The data provide observations on realizations of T_m. In addition, if T_p is completed before the realization t_m then we also observe the realization t_p, otherwise we merely observe that T_p exceeds t_m. The individuals who are observed to receive a treatment at a date t_p are a selected subset from the population under study. The most important reason for this is that the distribution of V_p among them does not equal the corresponding population distribution, because most individuals with high values of V_p have already had the treatment before. If V_p and V_m are dependent, then by implication the distribution of V_m among them does not equal the corresponding population distribution either. A second reason for why the individuals who are observed to receive a treatment at a date t_p are a selected subset is that, in order to *observe* the fact that treatment occurs at t_p, the individual should not have left the state of interest before t_p. Because of all this, the treatment effect cannot be inferred from a direct comparison of realized durations t_m of these individuals to the realized durations of other individuals. If the individuals with a treatment at t_p have relatively short durations then this can be for two reasons: *(1)* the individual treatment effect is positive, or *(2)* these individuals have relatively high values of V_m and would have left the state of interest relatively fast anyway. The second relation is called a *spurious* relation as it is merely due to the limited observability of the set of explanatory variables. This relation is also referred to as "selectivity". If V_m and V_p are independent then $\text{I}(t > t_p)$ is an "ordinary" exogenous time-varying covariate for T_m, and one may infer the treatment effect from a univariate duration analysis based on the distribution of $T_m|T_p = t_p, V_m$ mixed over the distribution of V_m. However, in general there is no reason to assume independence of V_m and V_p, and if this possible dependence is ignored then inference on the treatment effect may lead to incorrect conclusions.

3 Opposite cases

3.1 The pure treatment effect case

Our empirical procedure to assess whether $\delta >< 1$ is based on the following idea. Consider the subset of individuals whose spells end at a given duration $T_m = t_m$. Note that T_m is always observed, whether $T_p < T_m$ or not, so any t_m can be chosen here. If treatment increases the exit rate (i.e. if $\delta > 1$) then a relatively large fraction of those who exit at t_m have been treated shortly before t_m. Thus, conditional on t_m, the rate at which treatment is given $\theta_p(t|T_m = t_m)$ will tend to increase shortly before $t = t_m$. Of course, it still remains to show that this cannot be explained by duration dependence or unobserved heterogeneity as well. To deal with this, we will compare aspects of $\theta_p(t|T_m = t_m)$ for different values of t_m. In Section 5 we discuss an alternative approach which compares aspects of $\theta_m(t|T_p = t_p)$ as a function of t, for different values of t_p, and at values of t that exceed the largest t_p. We argue that this approach is less attractive than the one based on $\theta_p(t|T_m = t_m)$.

We analyze the behavior of $\theta_p(t|T_m = t_m)$ (or, in short-hand notation, $\theta_p(t|t_m)$) in two opposite extreme cases. Case I concerns the "pure treatment effect" case, where δ may differ from 1 but a selection effect is absent. The latter amounts to the assumption that V_m and V_p are independent. Case II concerns the "pure selection" case, where there can be dependent heterogeneity but $\delta = 1$. In this subsection and the next we show that these cases can be distinguished by examining the dependence of $\theta_p(t|t_m)$ on t and t_m. This enables us to construct a simple empirical check on whether there is a positive treatment effect.

The present subsection deals with the first case,

Case I (Pure treatment effect). $V_m \perp\!\!\!\perp V_p$.

We denote the marginal distributions of V_m and V_p by G_m and G_p, respectively, so that in Case I, $G(v_m, v_p) = G_m(v_m)G_p(v_p)$. We are able to derive more elegant and slightly stronger results under a more stringent version of our Case I, denoted as Case Ia:

Case Ia (Pure treatment effect). G_m is degenerate.

For expositional reasons we start by considering an even simpler model version in which there is no duration dependence or unobserved heterogeneity. As noted above, we expect $\theta_p(t|t_m)$ to increase as a function of t if and only if $\delta > 1$. This can be confirmed easily for this model version. We write $\theta_p(t|V_p) = \widetilde{\lambda}_p$ and $\theta_m(t|t_p, V_m) = \widetilde{\lambda}_m \cdot \delta^{\mathrm{I}(t>t_p)}$. It is also useful to define $\delta^* := \widetilde{\lambda}_p + (1-\delta)\widetilde{\lambda}_m$. After some calculations, using Bayes' rule, it follows that

$$\theta_p(t|t_m) = \frac{\delta^* \delta \widetilde{\lambda}_p}{\delta \widetilde{\lambda}_p + (1-\delta)(\widetilde{\lambda}_m + \widetilde{\lambda}_p)e^{-\delta^*(t_m - t)}} \quad \text{with } t \in [0, t_m] \tag{2}$$

for the generic case in which $\delta^* \neq 0$. Note that numerator and denominator in the right-hand side of equation (2) both have the same sign as δ^*. For the special case in which $\delta^* = 0$ (which implies that $\delta > 1$) we obtain

$$\theta_p(t|t_m) = \frac{\delta\widetilde{\lambda}_p}{1+\delta\widetilde{\lambda}_p(t_m - t)} \quad \text{with } t \in [0, t_m] \tag{3}$$

As a result, $\theta_p(t|t_m)$ increases in t if and only if $\delta > 1$. This is true for all parameter values and for any t_m, and, given t_m, for any $t \in [0, t_m]$.[6] Moreover, $\delta > 1$ if and only if $\theta_p(t|t_m)$ decreases in t_m, since $\theta_p(t|t_m)$ only depends on the difference of t_m and t.

Now let us examine what happens if we allow for duration dependence, i.e. if λ_m and λ_p are allowed to depend on the corresponding elapsed duration. The shape of $\theta_p(t|t_m)$ as a function of t will reflect this. For example, if $\lambda_p(t)$ displays a spike at a certain value of t then $\theta_p(t|t_m)$ also displays a spike at t. This is true for any given value of $t_m > t$. This can be illustrated by the following identity which can be shown to hold by definition (using Bayes' rule) for any continuous-time bivariate duration model,

$$\theta_p(t|t_m) = \theta_p(t) \cdot \lim_{dt \downarrow 0} \frac{\Pr(T_m \in [t_m, t_m + dt)|T_p = t)}{\Pr(T_m \in [t_m, t_m + dt)|T_p \geq t)} \tag{4}$$

where $\theta_p(t)$ is the hazard rate of the marginal distribution of T_p. Under Case I, T_p is an exogenous determinant of T_m, so the value of the duration dependence term λ_p at t enters the right-hand side only by way of its effect on $\theta_p(t)$. The latter is independent of t_m and acts multiplicatively on $\theta_p(t|t_m)$, so we can get rid of the effect of the duration dependence term λ_p on $\theta_p(t - t_m)$ by comparing $\partial \log \theta_p(t|t_m)/\partial t$ for different values of t_m. Basically, $\theta_p(t|t_m)$ may increase shortly before a given t_m because of duration dependence, but this can be corrected for by comparing the curve to curves corresponding to larger values of t_m. More precisely, it can be shown that the derivative of $\partial \log \theta_p(t|t_m)/\partial t$ with respect to t_m, which is of course the cross-derivative $\partial^2 \log \theta_p(t|t_m)/\partial t \partial t_m$, always has the same sign as $1 - \delta$ for $t < t_m$. Thus, without heterogeneity, the cross-derivative of $\log \theta_p(t|t_m)$ at a point $t < t_m$ provides sufficient information to infer the sign of $\delta - 1$. If $\delta > 1$ then $\log \theta_p(t|t'_m)$ increases in t relative to $\log \theta_p(t|t_m)$, for each $0 < t < t'_m < t_m$. It can also be shown that, without heterogeneity, $\partial \log \theta_p(t|t_m)/\partial t_m$ has the same sign as $1 - \delta$ for $t < t_m$. This reflects the fact that $\delta < 1$ ($\delta > 1$) generates a

[6] In fact, $\delta > 1$ if and only if $\theta_p(t|t_m)$ is convex in t on $(0, t_m)$.

negative (positive) association between T_m and T_p, which translates in a positive (negative) relation between t_m and $\log\theta_p(t|t_m)$. So, if $\delta > 1$ then the graph of $\log\theta_p(t|t_m)$ as a function of t is lower for higher values of t_m.

Now let us examine what happens if we also allow for independent heterogeneity of V_p and V_m. This generates (additional sources of) negative duration dependence in the observable hazard rates due to "dynamic selection". This selection mechanism should not be confused with selectivity in the treatment assignment. Dynamic selection concerns the fact that, as time proceeds, the composition of survivors shifts towards subjects with unfavorable unobserved heterogeneity values. Intuitively, this dynamic selection does not lead to duration dependence patterns in the observable hazard rates that vary substantially with t_m. Indeed, we prove that the comparison across various t_m of $\partial\log\theta_p(t|t_m)/\partial t$ as a function of t also effectively deals with the heterogeneity effects. The expected value of V_p conditional on survival up to time t is decreasing in t, but this effect is the same for every t_m. The heterogeneity in V_m complicates the derivations but does not lead to substantially weaker results. Concerning the effect of t_m on the level of $\theta_p(t|t_m)$, we find that this is also still informative on the sign of $\delta - 1$.

Specifically,

Proposition 1. *In Case I,*

$$\lim_{t\uparrow t'_m}\log\left[\frac{\theta_p(t|t_m)}{\theta_p(t|t'_m)}\right]\quad\begin{cases} > 0 \textit{ if } \delta < 1\\ = 0 \textit{ if } \delta = 1\\ < 0 \textit{ if } \delta > 1\end{cases}$$

and

$$\lim_{t\uparrow t'_m}\left[\frac{\partial\log\theta_p(t|t_m)}{\partial t}-\frac{\partial\log\theta_p(t|t'_m)}{\partial t}\right]\quad\begin{cases} > 0 \textit{ if } \delta < 1\\ = 0 \textit{ if } \delta = 1\\ < 0 \textit{ if } \delta > 1\end{cases}$$

for all $0 < t'_m < t_m$.

Proof. See Appendix 1, and Appendix 1.2 in particular. □

Suppose that $\delta > 1$, and let $t'_m < t_m$. The graph of $\log\theta_p(t|t'_m)$ as a function of $t \in (0, t'_m)$ lies above that of $\log\theta_p(t|t_m)$, at least for t just below t'_m. Moreover, the graphs of the log hazard rates diverge as functions of t, at least for t just below t'_m. Somewhat loosely, $\log\theta_p(t|t'_m)$ increases relatively strongly in t if t'_m is small. For $\delta < 1$, the ordering is reversed.[7]

[7]Instead of examining differences between expressions for t_m and t'_m in Proposition 1, we could as well have focused on derivatives with respect to t_m at a value t_m. However, the former is more general. Proposition 1 allows for a global comparison of the expressions for various values of t_m and t'_m.

In Appendix 1.3 we show by way of an example that there are distributions of V_m for which the results in Proposition 1 can not be generalized to hold for all $t \in (0, t'_m)$. However, in the more restrictive Case Ia,

Proposition 2. *In Case Ia,*

$$\frac{\partial \log \theta_p(t|t_m)}{\partial t_m} \begin{cases} > 0 \text{ if } \delta < 1 \\ = 0 \text{ if } \delta = 1 \\ < 0 \text{ if } \delta > 1 \end{cases}$$

and

$$\frac{\partial^2 \log \theta_p(t|t_m)}{\partial t \partial t_m} \begin{cases} > 0 \text{ if } \delta < 1 \\ = 0 \text{ if } \delta = 1 \\ < 0 \text{ if } \delta > 1 \end{cases}$$

for all $0 < t < t_m$.

Proof. See Appendix 1.4. □

We illustrate these results by way of a number of examples. First, consider the special case of Case Ia where there is no heterogeneity and no duration dependence (see equations (2) and (3)). We normalize time by fixing $\widetilde{\lambda}_p \equiv 1$. Figures 1 and 2 plot $\log \theta_p(t|t_m)$ for $\widetilde{\lambda}_m = 4$, and $\delta = 0.8$ and $\delta = 1.2$, respectively. In either figure, each curve corresponds to $\log \theta_p(t|t_m)$ on $(0, t_m)$ for a single value of t_m ($t_m = 0.5, 1, \ldots, 10$). As the end points of the intervals on which $\log \theta_p(t|t_m)$ is plotted coincide with t_m, we omit the legend from the graph. It is easy to see that the graphs are consistent with Proposition 2.

Figures 3 and 4 provide similar graphs for the case in which G_m and G_p are both unit exponential (so $G_i(v_i) = 1 - \mathrm{e}^{-v_i}$). The results are basically the same as in Figures 1 and 2, but the effects are attenuated at higher levels of t_m due to the effects of the unobserved heterogeneity.

3.2 The pure selection case

Before we provide a full characterization of Case II, we introduce some terminology and notation. For positive functions k and l, let $k(y) \overset{y \downarrow a}{\sim} l(y)$ denote $\lim_{y \downarrow a} k(y)/l(y) = 1$ (see Feller, 1971).

Definition 1. A positive function L defined on $(0, \infty)$ is *slowly varying* at 0 if $\lim_{y \downarrow 0} L(y\alpha)/L(y) = 1$ for every fixed $\alpha \in (0, \infty)$.

Definition 2. A positive function k defined on $(0,\infty)$ is *regularly varying with exponent* $-\infty < \rho < \infty$ at 0 if $k(y) \overset{y\downarrow 0}{\sim} y^{\rho}L(y)$ for a function L that is slowly varying at 0.

Case II is defined by

Case II (Pure selection). $\delta \equiv 1$. Furthermore, the joint distribution G of (V_m, V_p) is such that $V_p = V_m^{\tau}$ for some $\tau \geq 0$, where V_m has an absolutely continuous distribution function F. The corresponding density f is regularly varying at 0 with exponent $-1 < \sigma < \infty$.

This definition imposes that the relation between V_m and V_p is deterministic and nonnegative with one unknown parameter, that V_m and V_p are continuously distributed, and that the densities have a regular left-hand tail. We will show that this actually covers a very wide range of flexible specifications. In Appendix 2 we consider more general relations $V_p = h(V_m)$ between V_m and V_p, where h is not required to be non-decreasing, and we show that the results below can be generalized to include such relations. In Appendix 2 we also allow for certain classes of discrete distributions. For expositional convenience, we do not present those general results here. We briefly turn to them at the end of this subsection, and we also discuss results for other distributions there.

We proceed to discuss the notable features of Case II in some more detail. First, consider the relation between V_m and V_p. It can be generalized to $V_p = c \cdot V_m^{\tau}$ with $c > 0$, but this is equivalent to changing the scale of the baseline hazards. The deterministic relation effectively imposes that V_m and V_p satisfy a one-factor loading specification. This specification is best known in the representation $V_m = \exp(c_m\omega)$, $V_p = \exp(c_p\omega)$, where ω is a random variable with suitably normalized moments (note that we take $\tau := c_p/c_m$). The one-factor loading specification for unobserved heterogeneity in multivariate duration models was introduced by Flinn and Heckman (1982) and has become extremely popular in empirical research (see Van den Berg, 2001, for an overview).

Another notable feature is that the relation between V_m and V_p is assumed to be nonnegative (specifically, the correlation between $\log V_m$ and $\log V_p$ is zero or one). This is not so restrictive as it may seem. Often, the economically interesting issue is whether $\delta > 1$ or $\delta = 1$. This is because an intervention is often intended to increase the exit rate out of an undesirable state. If we are interested in testing $\delta > 1$ versus $\delta = 1$, the main problem is that both $\delta > 1$ and positively related unobservables induce a positive association between T_m and T_p. As such, the restriction to positively related unobservables does not solve the selectivity problem, and leaves a non-trivial identification task.

Yet another notable feature concerns the restriction to densities that are regularly varying at zero. This basically requires that the density just above 0 should not be too thin and should not have an irregular shape. Examples include densities with finite positive limits at 0 (like exponential and uniform densities) and a wide variety of densities truncated at 0 (like truncated normal densities). It also includes densities that converge to 0 at a polynomial rate as $v \downarrow 0$, like gamma densities and densities kv^k on $(0, 1)$, for $k > 0$. Furthermore, it includes certain densities that have infinite limits at 0 (see also Feller, 1971). The rationale behind this requirement is that it facilitates the analysis of the observed hazard rates at high durations. As time proceeds, the individuals with high values of the unobserved heterogeneity terms leave the state of interest, and the shape of F near the lower bound of its support determines "how much" heterogeneity is left among the survivors. Somewhat loosely, one may say that if the density of V_m just above 0 is not very thin then the heterogeneity among survivors is sufficient to affect the observed hazard rates at high durations. It should be noted, however, that our method of inference often also works well if the distribution of V_m is not covered by Case II (or by the more general case discussed in Appendix 2), i.e, if the density just above 0 is actually very thin (see below).

The condition on G_m in Case II should not be confused with a finite mean condition like $\mathrm{E}(V_m) < \infty$ that is typically imposed on the unobserved heterogeneity term in non-parametric analyses of the mixed proportional hazard model (e.g., Elbers and Ridder, 1982). Our condition deals with the left-hand tail of G_m, whereas the finite mean assumption from the non-parametric mixed proportional hazard model literature deals with the right-hand tail. We do not make assumptions on the right-hand tail or the moments of V_m.

We aim to contrast the behavior of $\theta_p(t|t_m)$ in Case II to the behavior in Case I. It is easy to see that in the model laid out by equations (1) with $\delta = 1$,

$$\theta_p(t|t_m) = \lambda_p(t)\ \mathrm{E}(V_p|T_p \geq t,\ T_m = t_m) \tag{5}$$

regardless of whether Case II holds or not. Let G have density g. By Bayes' rule, the distribution of $V_p|T_p \geq t,\ T_m = t_m$ has density

$$\frac{\int_0^\infty v_m\ \mathrm{e}^{-\Lambda_p(t)v_p - \Lambda_m(t_m)v_m}\ g(v_m, v_p)\ dv_m}{\int_0^\infty \int_0^\infty v_m\ \mathrm{e}^{-\Lambda_p(t)v_s - \Lambda_m(t_m)v_m}\ dG(v_m, v_s)} \tag{6}$$

If t increases then there are two effects on $\theta_p(t|t_m)$. First, there is the proportional duration dependence effect $\lambda_p(t)$. Secondly, there is the effect on the distribution of unobserved heterogeneity V_p: a large value of t makes smaller values of V_p more likely, and this reduces the level of the hazard rate aggregated over V_p.

If t_m increases then there is only one effect on $\theta_p(t_m|t)$, and this works by way of the distribution of V_p. A large value of t_m makes smaller values of V_m more likely. If V_m and V_p are positively related (as in Case II), then this makes smaller values of V_p more likely. As a result, we expect t_m to have a negative effect on the level of $\theta_p(t_m|t)$. This sign is the same as in Case I, so it seems that the effect of t_m on $\theta_p(t_m|t)$ cannot be used to distinguish between Case I and (positively related unobservables in) Case II.

Now consider the interaction between t and t_m in $\log \theta_p(t|t_m)$. By analogy to Case I, the evaluation of $\partial \log \theta_p(t|t_m)/\partial t$ at various values of t_m serves to eliminate the proportional duration dependence effect on $\theta_p(t_m|t)$ (see also equation (5)). It turns out that the interaction can indeed be used to distinguish between Case I and Case II,

Proposition 3. *In Case II, for each $t > 0$,*

$$\lim_{t_m \to \infty} \frac{\Lambda_m(t_m)}{\lambda_m(t_m)} \frac{\partial \log \theta_p(t|t_m)}{\partial t_m} = -\tau \leq 0,$$

and

$$\lim_{t_m \to \infty} \frac{\Lambda_m(t_m)^{\tau+1}}{\lambda_m(t_m)} \frac{\partial^2 \log \theta_p(t|t_m)}{\partial t \partial t_m} =$$

$$\tau \lambda_p(t) \left[\frac{\Gamma(\sigma+2+2\tau)}{\Gamma(\sigma+2+\tau)} - \frac{\Gamma(\sigma+2+\tau)}{\Gamma(\sigma+2)} \right] \begin{cases} > 0 \ \textit{if} \ \tau > 0 \\ = 0 \ \textit{if} \ \tau = 0 \end{cases}$$

Proof. See Appendix 2, which includes the proof of the proposition for an extension of Case II. □

In words, suppose that there is positive selection (i.e., individuals with a high treatment rate also have a high exit rate out of the current state). Then the graph of $\log \theta_p(t|t_m)$ is lower if t_m is larger, at least if the values of t_m are sufficiently large. Moreover, in that case, the graphs of the log hazard rates converge as functions of t.

Note that Proposition 3 examines derivatives in the limit as $t_m \to \infty$, whereas Proposition 1 for Case I examines derivatives at arbitrary t_m. This may make Proposition 3 look restrictive and less relevant. However, for reasons of smoothness, the derivatives in Case II with $\tau > 0$ generally have the limiting sign well before the limit is reached. Below we show in detail that in many instances the predictions in Proposition 3 hold true for every t_m, even when the distribution

G does not satisfy the description of Case II at all. In practice, to distinguish between Case I and Case II one would have to examine large values of t_m.

We provide intuition behind Proposition 3 by considering the special case where $\tau = 1$, i.e. $V_p = V$, $V_m = V$, and V has distribution function F concentrated on $(0, \infty)$. In that case, equation (5) reduces to $\theta_p(t|t_m) = \lambda_p(t)\,\mathrm{E}(V|T_p \geq t, T_m = t_m)$. Moreover, derivatives of $\log\theta_p(t|t_m)$ with respect to t and/or t_m can be expressed in terms of moments of $[V|T_m = t_m,\ T_p \geq t]$. The family of gamma distributions plays an important role in the intuition that we give. It is not difficult to see that if F is a gamma distribution with parameters r, a, i.e., if the density $f(v)$ equals

$$f(v) = \frac{a^r}{\Gamma(r)} v^{r-1} \mathrm{e}^{-av} \qquad \text{with} \quad a, r > 0$$

then the equations in Proposition 3 hold, with $\tau = 1$ and $\sigma := r - 1$, for *every* $t_m > 0$, i.e., not just in the limit as $t_m \to \infty$. Now suppose that F is, more generally, a distribution that satisfies Case II. We introduce short-hand notation

$$z := \Lambda_m(t_m) + \Lambda_p(t).$$

From the results in Appendix 2 it follows that the moments of $z \cdot [V|T_m = t_m, T_p \geq t]$ converge to the moments of a gamma distribution with parameters $\sigma + 1, 1$. This suggests that the distribution of $[V|T_m = t_m,\ T_p \geq t]$ becomes more and more similar to a gamma distribution, as time proceeds.

Hougaard, Harvald and Holm (1992) study the (non-causal) effect of conditioning on the realization of one duration variable on the log hazard rate for the other, if both share the same unknown unobserved heterogeneity term V. Specifically, they examine to what extent this effect depends on the type of distribution of V across duration pairs. For the gamma distribution, the effect increases in the time distance since the realization of the first duration variable (on which we condition). Let us translate this to our Case II model with $V_m = V_p = V$. We examine the effect of knowing that $T_m = t_m$ on the log hazard rate of T_p. Let V have a gamma distribution. If $T_p = t$ is close to t_m then the effect is relatively small, whereas if $T_p = t$ is much smaller than t_m then the effect is much larger. This implies that if one compares the function $\log\theta_p(t|t_m)$ for different values of t_m, then the effect of t_m is larger for small t. If V does not have a gamma distribution but satisfies Case II then this result holds as the t_m values become large. This explains the convergence of the graphs of the log hazard rates as given in Proposition 3. If $V_p = V_m^\tau$ then the moments of $[z^\tau \cdot V_p|T_m = t_m,\ T_p \geq t]$ converge to the moments of a generalized gamma distribution (see McDonald, 1984, for a description of this distribution), and a similar intuition can be given.

We now examine the special case with $V_m = V_p = V$ from another angle. By analogy to (6), it follows that $[V|T_m = t_m,\, T_p \geq t]$ has the distribution function $F_z(v) := \int_0^v dF_z(\eta)$ with

$$dF_z(v) := \frac{v\exp(-zv)dF(v)}{\int_0^\infty v\exp(-zv)dF(v)} \tag{7}$$

if $t_m > 0$ and $t > 0$. For convenience, we define $V_z := [V|T_m = t_m,\, T_p \geq t]$. Let $\mu(z)$ denote the expectation of the distribution F_z, so $\mu(z) := \mathrm{E}(V_z)$. Equation (5) then reduces to $\theta_p(t|t_m) = \lambda_p(t)\mu(z)$. We denote the normalized centralized moments of V_z by $\widetilde{\gamma}_i(z) := \mathrm{E}(V_z - \mu(z))^i/\mu(z)^i$. In addition, we denote the coefficient of variation by $\nu_2(z) := \sqrt{\widetilde{\gamma}_2(z)}$, and Cox and Oakes' (1984) standardized index of skewness by $\nu_3(z) := \widetilde{\gamma}_3(z)/\widetilde{\gamma}_2(z)^{3/2}$. Using the equations in Appendix 2.2, it is easy to verify that

$$\frac{\partial \log\theta_p(t|t_m)}{\partial t_m} = -\lambda_m(t_m)\mu(z)\widetilde{\gamma}_2(z)$$

and

$$\frac{\partial^2 \log\theta_p(t|t_m)}{\partial t\partial t_m} = \lambda_p(t)\lambda_m(t_m)\mu(z)^2\widetilde{\gamma}_2(z)^{3/2}\left[\nu_3(z) - \nu_2(z)\right],$$

Clearly, $\partial \log\theta_p(t|t_m)/\partial t_m < 0$. Note that this is true for *all* $t_m > 0$ if $V_m = V_p = V$, and that this is true regardless of whether the left-hand tail of the density has a regular shape, and, indeed, regardless of whether V is continuous. The interaction effect $\partial^2 \log\theta_p(t|t_m)/\partial t\partial t_m$ is positive if and only if

$$\nu_3(z) > \nu_2(z), \tag{8}$$

i.e. if the distribution of V_z is sufficiently skewed to the right relative to its dispersion.

Normalized variation and skewness indicators like ν_2 and ν_3 are frequently used to classify scale families of distributions (see e.g. Cox and Oakes, 1984). The inequality (8) is satisfied by all log-logistic, log-normal, and gamma distributions, and by some distributions in the Weibull class. Intuitively, this result is not surprising, as these distributions all resemble gamma distributions, and we know that the interaction effect is positive for all t_m if V has a gamma distribution. Note however that the inequality (8) concerns the distribution F_z of $V_z := [V|T_m = t_m,\, T_p \geq t]$. If F_z belongs to a well-known family of distributions then the underlying distribution F of V does not necessarily belong to a well-known family, and vice versa. The exception is when F is a gamma distribution with parameters r, a, since then F_z also is a gamma distribution with parameters $r+1, a+z$, and $\nu_2(z) = 1/\sqrt{r}$, $\nu_3(z) = 2/\sqrt{r}$, and $\nu_3(z)/\nu_2(z) = 2$ for all $z \geq 0$. A special case of this is when V has an exponential distribution: then $r = 1$, and therefore $\nu_2(z) = 1$ and $\nu_3(z) = 2$.

The "skewness" condition in equation (8) which ensures the desired sign of the interaction between t and t_m in $\log \theta(t|t_m)$ does not refer to the "regularity" condition on $f(v)$ close to $v = 0$ in the characterization of Case II and its generalization in Appendix 2. For example, if V has a log-normal distribution and z is close to zero then the "skewness" condition is satisfied whereas the "regularity" condition is not, as log-normal densities are not regularly varying at zero (we return to log-normal heterogeneity distributions below). This suggests that the "regularity" condition in Case II is by no means necessary to obtain the desired interaction sign.

Now let us examine discrete distributions for V_m with a finite number of positive points of support. Such distributions have *zero* probability mass right next to 0, so they are not covered by Case II or its generalization in Appendix 2. As time proceeds, individuals with V_m exceeding the smallest point of support leave the state of interest rather quickly, and at very large t the survivors are virtually homogeneous. In such a case, for sufficiently large t_m, the conditional hazard rate $\theta_p(t|t_m)$ behaves as in the model without unobserved heterogeneity, so the interaction effect $\partial^2 \log \theta_p(t|t_m)/\partial t \partial t_m$ converges to zero.[8]

In Appendix 2 we consider more general relations $V_p = h(V_m)$ between V_m and V_p, where h is not required to be non-decreasing. Basically, if T_m and T_p are negatively related then the effect of t_m on $\log \theta_p(t|t_m)$ is positive. It turns out that, in addition, the limiting interaction effect may be positive or zero. This means that on the basis of these signs it cannot be distinguished from Case I with $\delta \leq 1$.

We illustrate the results of this subsection by plotting $\log \theta_p(t|t_m)$ in some specific examples. Like in Subsection 3.1, we exclude duration dependence and we take $\lambda_p \equiv 1$ and $\lambda_m \equiv 4$. Figure 5 plots $\log \theta_p(t|t_m)$ for the case that F is unit exponential and $V_m = V_p$. This is a case in which there is a positive association between T_m and T_p, which is potentially confused with a positive effect of treatment ($\delta > 1$). We indeed find a negative effect of t_m on the level of $\log \theta_p(t|t_m)$ for given t, as in Figures 2 and 4 for Case I. However, consistent with

[8]If V_m has a continuous distribution F concentrated on $(\underline{v}, \infty)$, for some $\underline{v} \in (0, \infty)$, and where $F_0(v_m) := F(v_m + \underline{v})$ satisfies Case II, then our results can be translated to this case by using properties of Laplace transforms of translated random variables. We do not pursue this further here.

the propositions above, the interaction effect is now reversed relative to Case I. So, we are able to distinguish between these cases by exploiting the interaction effect. It should be noted that the effects of varying t_m disappear quite rapidly with increasing t.

In this example, the results hold for all t, t_m, so we do not need to restrict attention to large values of t_m only. Figure 6 provides an example where it is necessary to restrict attention to large t_m. In this example, $V_m = V_p = V$, and F is a left-skewed beta-distribution on $(0,1)$ ($F(v) = v^{11}$, which satisfies Case II). At all values of t_m, the log hazard $\log \theta_p(t|t_m)$ decreases in t_m. However, the sign of the interaction effect may lead to confusion with $\delta > 1$ in Case I (see Figures 2 and 4) if one examines small values of t_m. In particular, for small values of t_m, the change in the slope of $\log \theta_p(t|t_m)$ with t_m resembles Case I with $\delta > 1$. However, this changes rapidly with increasing t_m.

Figure 7 shows that the distinction between treatment effect and selectivity may also work well for heterogeneity distributions that are excluded under Case II and its generalization in Appendix 2. It plots $\log \theta_p(t|t_m)$ for the case that F is a standard log-normal distribution truncated from above at 2, and $V_m = V_p$. The resulting graph is not much different from that in Figure 5.

If $V_p = -\log(V_m)$, then T_m and T_p are negatively related. With F being unit exponential truncated from above at 1, this case is in the domain of the more general Proposition 4 in Appendix 2. As is to be expected, we find a positive effect of t_m on $\log \theta_p(t|t_m)$ (see Figure 8). We also find a positive interaction effect. This means that in terms of these signs it cannot be distinguished from Case I with $\delta < 1$ (see Figures 1 and 3). It can however be distinguished from Case I with $\delta > 1$.

4 Inference of a causal effect

4.1 A graphical check

The results of the previous section enable the construction of an informal test on a positive treatment effect. The basic procedure is as follows. First, choose some large values of t_m. Secondly, draw $\log \theta_p(t|t_m)$ as a function of t for each of these values of t_m. If the line is higher for the smaller t_m (in particular at durations t just below this t_m) then this means that there is a positive treatment effect and/or there are positively related unobserved determinants of T_m and T_p. If, in addition to this, the lines diverge, then there must be a positive treatment effect. If they converge then there must be positively related unobserved determinants. If they diverge then there may also be positively related unobserved determinants, but the selection effect is dominated by the positive treatment effect. Similarly, if the lines converge then there may also be a positive treatment effect but this is dominated by the selection effect. If the lines are parallel then both effects are present.

We now briefly compare our method of inference to the difference-in-differences method of inference on treatment effects in panel data. In both approaches, the data and model have time dimensions, and the treatment effect works from a specific point of time onwards, whereas the selection effect works at all points of time in a more permanent way. In both approaches, the inference focuses on interaction effects in the data, and one needs to make separability assumptions that rule out certain interaction effects of the determinants of the individual outcome of interest. In particular, in panel data models, additivity of the treatment effect, the unobserved heterogeneity, and the residual error term in the individual outcome equation is crucial. In our approach, additivity of the determinants of the individual log outcome hazard rate $\log \theta_m(t|t_p, V_m)$ is crucial. These separability assumptions at the individual level enable an empirical distinction between the treatment effect, that works from a specific point of time, and the selection effect that works at all points of time. Observed covariates do not play an important role in either case.

However, our approach is more involved, for the reason that we essentially only observe one outcome per individual (which implies that the composition in terms of unobserved heterogeneity terms changes over time). As a result, we cannot straightforwardly apply difference-in-differences. The individual-specific unobserved heterogeneity terms cannot be treated as incidental parameters or fixed effects but are treated as realizations of random variables, and our examination of interaction effects involves a comparison over time of different individuals. The prices to be paid (relative to the panel data approach) are that the treatment assignment process has to be specified, that we have to assume that the individual treatment effect is constant after imposition, and that we are only able to make inference on the sign of the treatment effect.

Nevertheless, our results highlight the usefulness of the information in the timing of events to assess the treatment effect on a transition rate. This information is discarded in a binary treatment framework. Intuitively, if treatment and outcome are typically realized very quickly after each other, no matter what the values of the other observed outcome determinants are, then this is evidence of a positive causal treatment effect on the individual transition rate. The selection effect does not give rise to the same type of quick succession of events.

We now turn to some practical issues that arise when implementing our informal test. To plot the conditional log hazard rate $\log \theta_p(t|t_m)$ for given t_m we have to select the subsample of individuals with realization $T_m = t_m$. In fact, to obtain a positive subsample size, we have to select individuals with realizations in an interval around t_m, say (a, b). It is straightforward to nonparametrically estimate and plot $\theta_p(t|t_m)$ at values $t < a$. If $b \downarrow a$ then this estimator converges to a consistent estimator of the underlying hazard rate. Some care has to be taken to estimate the terminal value $\lim_{t \uparrow t_m} \theta_p(t|t_m)$, though. The observed empirical equivalent of $\Pr(a \le T_p < T_m | a \le T_m < b, T_p \ge a) \; / \; (b - a)$ is not a consistent estimator of this, because, basically, it ignores the probability that the latent variable T_p is realized in the interval (T_m, b). In fact, it can be shown that

$$\lim_{b \downarrow a} \frac{\Pr(a \le T_p < T_m | a \le T_m < b, T_p \ge a)}{b - a} \quad = \quad \frac{1}{2} \cdot \theta_p(a | T_m = a)$$

(this holds for any bivariate distribution). As a result, $\lim_{t \uparrow t_m} \theta_p(t|t_m)$ can be estimated by the observed empirical equivalent of $2 \Pr(a \le T_p < T_m | a \le T_m < b, T_p \ge a)/(b-a)$. For sizeable intervals (a, b), and with relatively smooth duration dependence, this may actually underestimate $\lim_{t \uparrow t_m} \theta_p(t|t_m)$, for the reason that most realizations of T_m and T_p are at the lower end of the interval.

This highlights the fact that the intervals for the values of t_m should be small. At the same time, it is useful to have large subsamples of individuals in a given interval for t_m, because that makes the plotted hazard rates less prone to sampling error, By conditioning t_m to be in a small time interval, we effectively consider a small subsample. Note that, at the same time, the chosen values of t_m should be rather large.[9] These demands can only be reconciled if the data set is very large.

4.2 Examples

In general, recipients of unemployment benefits have to comply with minimum requirements concerning search behavior. Compliance is imperfectly monitored. If violation of the search rules is detected, then a punitive sanction is imposed, consisting of a benefits reduction, and entailing an increase of monitoring in the

[9] One may wonder whether a particular t_m is sufficiently large. This can be investigated by examining $\theta_m(t|T_p = t)/\theta_m(t|T_p > t)$ as a function of t, which is directly estimable from the raw data. In Case II, as $t \to \infty$, this converges to a constant, and this reflects the convergence of the partial derivatives in Proposition 3. In Case I, $\theta_m(t|T_p = t)/\theta_m(t|T_p > t) = \delta$, which is also constant. Somewhat loosely, one may therefore argue that t_m is sufficiently large if the plot of $\theta_m(t|T_p = t)/\theta_m(t|T_p > t)$ as a function of t is flat at $t = t_m$.

future. A sanction can thus be expected to affect the re-employment rate of the individual. Abbring, Van den Berg and Van Ours (1997) and Van den Berg, Van der Klaauw and Van Ours (2004) analyze the effect of, respectively, unemployment insurance and welfare sanctions on re-employment rates by estimating models that are extensions of the model in this paper.

In Abbring, Van den Berg and Van Ours (1997), the full sample of unemployment insurance recipients contains about 148,000 individuals. Only about 4300 of these are observed to receive a "sanction" treatment, while about 44% of the spells are right-censored. It turns out that this sample is sufficiently large to construct graphical checks that are robust with respect to the choice of (the intervals of) the t_m values. (The graphical analysis does not stratify on X.) The graphical check provides strong evidence for the presence of a positive sanction effect on the individual transition rate into employment.[10] Estimation of the full bivariate duration model results in an estimate of δ that is significantly larger than one and in significant positively related unobserved heterogeneity. The latter suggests that both reasons for a positive relation between T_m and T_p given X are present, and that the causal treatment effect dominates in the interaction effect of t and t_m in $\log \theta_p(t|t_m)$.

In Van den Berg, Van der Klaauw and Van Ours (2004), the full sample of welfare recipients contains about 8,000 individuals. About 1100 of these are observed to receive a "sanction" treatment, and about 60% of the spells are right-censored. This sample is too limited to construct sensible graphical checks. The plotted shapes of the hazard rates are erratic and strongly dependent on the choice of (the intervals of) the t_m values.[11] The estimate of δ obtained by estimation of the full bivariate duration model is significantly larger than one, and there is significant positively related unobserved heterogeneity. We tackle the problems with small sample sizes in Subsection 4.3 below.

Richardson and Van den Berg (2002) study the effect of participation in a vocational training program by unemployed workers on their transition rate to work. Individuals spend on average 6 months in training. Presumably, job search continues during the training. The data show that a large fraction of individuals move to work in the days following exit from the training program. This can not be captured by the model of Section 2. Whether one halts the time clock during the training or not, the transition rate to work depends on the time since the start of the treatment. If one halts the clock then the figure produced by the graphical check merely displays enormous peaks at the moment of exit to work.

[10] See Abbring, Van den Berg and Van Ours (1997) for the figure.

[11] This is not discussed in Van den Berg, Van der Klaauw and Van Ours (2004).

If one does not then then these peaks occur 6 months before exit to work. The graphical check is better suited for situations in which the treatment is permanent and the treatment is either instantaneous or is time consuming but takes place in "quarantine" (i.e., all activities regarding the outcome of interest other than the treatment are put on hold). However, it is not difficult to envisage modifications of the graphical check that allow for a positive joint occurrence of treatment and outcome and for a treatment effect that decreases as a function of the time since treatment. We leave this for future work.

4.3 Estimation of an auxiliary duration model

To get around the practical problems of discretization and sample size requirements, one may formalize our preferred graphical procedure by estimating an ad hoc descriptive specification for $\theta_p(t|t_m)$ and test the signs of the estimated effects of t and t_m (and their interaction) on $\log \theta_p(t|t_m)$. This way, we may simultaneously use the information on all individuals in the sample who are observed to realize T_m, including those who are not observed to receive a treatment. In addition, we may allow the parameters of the ad hoc specification to depend on X. The specification is only relevant for $T_p < t_m$.

For example, we may specify for $t < t_m$ that

$$\theta_p(t|x, t_m) = \exp(x'\beta_1 + \beta_2 \log t_m)\ \alpha(t_m)\ t^{\alpha(t_m)-1} \tag{9}$$

$$\text{with} \quad \alpha(t_m) = \exp(\alpha_0 + \alpha_1 \log t_m)$$

It should be stressed that this specification merely summarizes the data. The parameters should not be given causal interpretations. Rather, they represent patterns across individuals, between X and T_m on the one hand and T_p on the other.

The above specification is a univariate duration model, where the hazard rate $\theta_p(t|t_m)$ follows a Weibull specification, with the duration dependence parameter α being dependent on the "explanatory" variable t_m. It is straightforward to derive the likelihood function corresponding to i.i.d. observations of $T_p|X = x, T_m = t_m$, where T_p may be right-censored. Note that the censoring variable is the "explanatory" variable T_m. One may also restrict attention to relatively large values of T_p (i.e. truncate T_p from below). The parameter estimates can be used to test the signs of the effects of t and t_m. Notably, the cross-derivative of $\log \theta_p(t|t_m)$ with respect to t and t_m is positive (negative) if

$\alpha_1 > 0$ (if $\alpha_1 < 0$).[12]

Note that we have thus effectively reduced the dimension of the duration analysis from 2 (in the full model) to 1. Of course, the specification (9) does not follow from the full model. It is desirable that the sign of the probability limit of the estimator of α_1 equals the sign of the cross-derivative in the underlying full model. We feel that the derivation of results on this is beyond the scope of the paper.

We applied the above univariate approach to the data on sanctions of welfare recipients. It turns out that $\theta_p(t|t_m)$ is estimated to decrease with t_m while the estimate of α_1 is insignificantly different from zero.[13] This implies that there is a positive treatment effect as well as a positive selection effect, which is in agreement to the estimates for the corresponding full bivariate model.

5 An alternative graphical check

We now discuss the extent to which $\theta_m(t|T_p = t_p)$ (or, in short-hand notation, $\theta_m(t|t_p)$) can be used for an alternative graphical check on the treatment effect. The idea is that the log of this hazard rate can be plotted for say two different values of t_p, and that the resulting lines can be compared for t exceeding the largest of the two values of t_p. In fact, in Case II, the hazard rates $\theta_m(t|t_p)$ and $\theta_p(t|t_m)$ are closely related. This is easy to see if $V_m \equiv V_p$. For example, we then have that

$$\frac{\theta_p(t|t_{m,1})}{\theta_p(t|t_{m,2})} = \frac{\theta_m(t_{m,1}|t)}{\theta_m(t_{m,2}|t)} \quad \text{and} \quad \frac{\partial^2 \log \theta_p(t|t_m)}{\partial t \partial t_m} = \frac{\partial^2 \log \theta_m(t_m|t)}{\partial t \partial t_m}$$

for all possible realizations $t, t_m, t_{m,1}, t_{m,2}$.[14]

[12] One may of course adopt alternative specifications, like $\theta_p(t|x, t_m) = \exp(x'\beta_1 + \beta_2 \log t_m + \alpha_0 \log t + \alpha_1 (\log t) \cdot (\log t_m))$, which is also a Weibull specification. Here as well, the cross-derivative is positive (negative) if $\alpha_1 > 0$ (if $\alpha_1 < 0$). More general specifications may allow the cross-derivative to vary with t, t_m and x in a flexible way, but then the simplicity appeal of the estimation procedure is lost.

[13] These results are not reported in Van den Berg, Van der Klaauw and Van Ours (2004).

[14] In fact, in Case II, the properties of the hazard rates are closely related to the properties of $\theta_m(t|T_p = t_p)/\theta_m(t|T_p > t)$ as a function of t for different t_p. The latter is often used to study the way in which the dependence of two duration variables changes over time if the dependence runs by way of a common unobserved heterogeneity term (see e.g. Hougaard, Harvald and Holm, 1992, and Yashin and Iachine, 1999). We conjecture that this can also be used for a graphical check for our purposes, but we do not pursue this further.

It is easy to show that with $V_m \equiv V_p$ in Case II, the hazard $\theta_m(t|t_p)$ as seen as a function of t is always higher for smaller t_p. It can also be shown that in Case I with $\delta > 1$, $\theta_m(t|t_p)$ is always *lower* for smaller t_p. The latter result can be understood as follows: if there is a positive treatment effect at say $t_p = 1$ then the individuals with high values of V_m leave the state very quickly, so that the survivors at say $t_p = 2$ have smaller values of V_m than if the treatment would be given at $t_p = 2$. Note that in Case Ia, the hazard $\theta_m(t|t_p)$ as a function of t on (t_p, ∞) does not depend on t_p at all. If $V_m \neq V_p$ in Case II then we have to focus on high durations again.

This implies that a sensible check on the treatment effect can be based on $\theta_m(t|t_p)$, and, indeed, on the sign of the effect of t_p on this hazard rate. There are however a number of reasons to prefer the check based on $\theta_p(t|t_m)$. Taken in isolation, neither of these reasons may be sufficiently convincing, but taken together we feel that they are.[15] First of all, consider the result that $\theta_m(t|t_p)$ increases with t_p in Case I with $\delta > 1$. The above-mentioned intuition behind it makes clear that this is critically dependent on the proportionality of the treatment effect and the unobserved heterogeneity term in the individual hazard rate $\theta_m(t|t_p, V_m)$. The shape of $\theta_p(t|t_m)$ in Case I is likely to be less sensitive to this assumption. For example, the sign results on $\theta_p(t|t_m)$ in Case I do not depend on whether there is unobserved heterogeneity or not, whereas the sign results on $\theta_m(t|t_p)$ in Case I do. Also, suppose that δ is a not constant but instead decreases slightly with the realization of T_p, while $\delta(t_p) > 1$ everywhere. Then in Case Ia, $\theta_m(t|t_p)$ is always *higher* for smaller t_p, so that the check based on $\theta_m(t|t_p)$ leads to the wrong conclusion, whereas the check based on $\theta_p(t|t_m)$ does not.

Another reason concerns the fact that with $\theta_m(t|t_p)$ it is difficult to distinguish between $\delta < 1$ and positively related unobserved heterogeneity. In both of these cases, $\log \theta_m(t|t_p)$ may be larger for small t_p, and in both cases the interaction term may be positive, so that the lines corresponding to different t_p converge to each other. Yet another reason concerns the fact that the check based on $\theta_p(t|t_m)$ also detects a treatment effect in structural nested failure time models (see Robins, 1998, and Keiding, 1999) whereas the check based on $\theta_m(t|t_p)$ does not. Structural nested failure time models are popular in biostatistics as a framework to study the effect of treatments over time on duration variables. In our terminology, it is assumed that V_p is degenerate, so that all systematic deter-

[15] One may wonder whether a graphical check based on $\theta_m(t|t_p)$ can be designed for multiple-spell data, using variation in t_p across spells for given individuals. However, there is often insufficient information for such a graphical approach. The partial likelihood approach in Abbring and Van den Berg (2003) formalizes the underlying idea.

minants of whether an individual gets a treatment at a duration t_p are known. This implies that selectivity due to related unobserved heterogeneity is absent. The models do however allow for time-varying confounders, which are basically time-varying explanatory variables for θ_m that may depend on the (moment of) treatment.[16]

6 Conclusion

In this paper, we have developed a simple procedure to infer the presence of a positive treatment effect on a transition rate. The procedure can be implemented as a graphical check. In this check, one has to condition on the moment of exit and examine what happens before that (rather than condition on the moment of treatment and examines what happens after that). The check focuses on the shape of the rate at which a treatment is given conditional on the moment of exit, and as such it is easy to use. In practical cases where sample sizes are too small to rely on the graphical check, it can be replaced by a formalized version in which the key properties of the shape are represented by parameters in an ad hoc auxiliary duration model. Our method demonstrates that variation in the duration until treatment relative to the duration until the outcome of interest conveys useful information on the causal treatment effect in the presence of selection effects. If treatment and outcome are typically realized very quickly after each other, no matter what the values of the other outcome determinants are, then this is taken as evidence of a positive causal treatment effect. The selection effect does not give rise to the same type of quick succession of events.

Some topics for further research remain. The results on the graphical procedure are derived conditional on the assumption that the treatment effect, the unobserved heterogeneity term, and the duration dependence term act multiplicatively on the hazard rates. Concerning the unobserved heterogeneity term,

[16]Yet another reason for preferring the check based on $\theta_p(t|t_m)$ follows from the fact that we need the durations t_m and/or t_p to be large. Consider again for convenience the special case that $V_m \equiv V_p$. Both in $\theta_p(t_p|t_m)$ and in $\theta_m(t_m|t_p)$ the relevant unobserved heterogeneity distribution is affected by a multiplicative term $\exp(-(\Lambda_m(t_m) + \Lambda_p(t_p))v)$ (this can be seen from equation (7) for $\theta_p(t|t_m)$), and from Appendix 2 it follows that $\Lambda_m(t_m) + \Lambda_p(t_p)$ needs to be large. Typically, in practical applications, $\Lambda_m(t) >> \Lambda_p(t)$ for all sufficiently large t. Also, in practice, both T_m and T_p are right-censored at a common value, say $\mathcal{T}$. Suppose that $\mathcal{T}$ is sufficiently large for the "large duration approximation" to be valid at some points $t_m, t_p < \mathcal{T}$. We may plot $\theta_p(t|t_m)$ for a few t_m just below $\mathcal{T}$, and the resulting lines may be compared in an interval for t. We may also plot $\theta_m(t|t_p)$ for a few t_p below $\mathcal{T}$, but the interval in which these lines may be compared is much smaller, which is unattractive.

this is essential for reasons of tractability, as this allows us to rely on the powerful Laplace transformation theory. The proportionality of the treatment effect and the duration dependence term may however be violated in practice, and it remains to be seen to what extent our results are robust with respect to this.

References

Abbring, J.H. and G.J. van den Berg (2003), "The non-parametric identification of treatment effects in duration models", *Econometrica*, forthcoming.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (1997), "The effect of unemployment insurance sanctions on the transition rate from unemployment to employment", Working paper, Tinbergen Institute, Amsterdam.

Angrist, J.D. and A.B. Krueger (1999), "Empirical strategies in labor economics", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Blundell, R. and T. MaCurdy (1999), "Labor supply", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Bonnal, L., D. Fougère, and A. Sérandon (1997), "Evaluating the impact of French employment policies on individual labour market histories", *Review of Economic Studies*, **64**, 683–713.

Card, D. and D. Sullivan (1988), "Measuring the effect of subsidized training programs on movements in and out of employment", *Econometrica*, **56**, 497–530.

Cox, D.R. and D. Oakes (1984), *Analysis of Survival Data*, Chapman and Hall, London.

Elbers, C. and G. Ridder (1982), "True and spurious duration dependence: The identifiability of the proportional hazard model", *Review of Economic Studies*, **64**, 403–409.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, volume II, 2nd edition, Wiley, New York.

Flinn, C.J. and J.J. Heckman (1982), "Models for the analysis of labor force dynamics", in R. Basmann and G. Rhodes, editors, *Advances in Econometrics, volume 1*, JAI Press, Greenwich.

Freund, J.E. (1961), "A bivariate extension of the exponential distribution", *Journal of the American Statistical Association*, **56**, 971–977.

Gritz, R.M. (1993), "The impact of training on the frequency and duration of employment", *Journal of Econometrics*, **57**, 21–51.

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999), "The economics and econometrics of active labor market programs", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Hougaard, P., B. Harvald, and N.V. Holm (1992), "Assessment of dependence in the life times of twins", in J.P. Klein and P.K. Goel, editors, *Survival Analysis: State of the Art*, Kluwer Academic Publishers, Dordrecht.

Keiding, N. (1999), "Event history analysis and inference from observational epidemiology", *Statistics in Medicine*, **18**, 2353–2363.

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

Lillard, L.A. (1993), "Simultaneous equations for hazards", *Journal of Econometrics*, **56**, 189–217.

Lillard, L.A. and C.W.A. Panis (1996), "Marital status and mortality: The role of health", *Demography*, **33**, 313–327.

McDonald, J.B. (1984), "Some generalized functions for the size distribution of income", *Econometrica*, **52**, 647–663.

Richardson, K. and G.J. van den Berg (2002), "Swedish labor market training and the duration of unemployment", Working paper, IFAU, Uppsala.

Robins, J.M. (1998), "Structural nested failure time models", in P. Armitage and T. Colton, editors, *The Encyclopedia of Biostatistics*, John Wiley and Sons, Chichester.

Tsiatis, A. (1975), "A nonidentifiability aspect of the problem of competing risks", *Proceedings of the National Academy of Sciences*, **72**, 20–22.

Van den Berg, G.J. (2001), "Duration models: Specification, identification, and multiple durations", in J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, Volume V*, North Holland, Amsterdam.

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**, forthcoming.

Yashin, A.I. and I.A. Iachine (1999), "What difference does the dependence between durations make?", *Lifetime Data Analysis*, **5**, 5–22.

Appendix

Appendix 1 Results for Case I: pure treatment effect

Appendix 1.1 Characterization of the data

The results in Appendix 1.1 apply to all Cases. By analogy to Tsiatis (1975), it is not difficult to see that the observed joint distribution of $T_m, \mathrm{I}(T_m > T_p)$, and $T_p \cdot \mathrm{I}(T_m > T_p)$ is characterized by the functions

$$Q_m(t) := \Pr(T_m > t, T_p > T_m) \text{ and } \tag{10}$$

$$Q_p(t, t_p) := \Pr(T_m > t, T_p > t_p, T_m > T_p). \tag{11}$$

(see also Lancaster, 1990).

By differentiation of (10) and (11) we obtain, in notation to be explained below,

$$\frac{\partial Q_m(t)}{\partial t} = \lambda_m(t) \cdot \mathcal{L}_G^{(m)}(\Lambda_m(t), \Lambda_p(t)) \tag{12}$$

and

$$\frac{\partial Q_p(t, t_p)}{\partial t_p} = \lambda_p(t_p) \cdot \mathcal{L}_G^{(p)}([\Lambda_m(t_p) + \Delta(t|t_p, x)], \Lambda_p(t_p)), \tag{13}$$

for $t, t_p \in (0, \infty)$ and $t_p < t$. $\mathcal{L}_G$ is the bivariate Laplace transform of G,

$$\mathcal{L}_G(s_m, s_p) := \int_0^\infty \int_0^\infty \exp(-s_m v_m - s_p v_p) dG(v_m, v_p)$$

and $\mathcal{L}_G^{(i)}(s_m, s_p)$ $(i = m, p)$ is the partial derivative of $\mathcal{L}_G(s_m, s_p)$ with respect to s_i.

Appendix 1.2 Proof of Proposition 1

Expressions for $\theta_p(t|t_m)$.

Note that $G(v_m, v_p) = G_m(v_m)G_p(v_p)$ implies that $\mathcal{L}(s,t) = \mathcal{L}_m(s)\mathcal{L}_p(t)$, where $\mathcal{L}_m$ and $\mathcal{L}_p$ are the Laplace transforms of respectively G_m and G_p. Let $M_\delta(t, t_m) := \Lambda_m(t) + \Delta(t_m, t) = (1-\delta)\Lambda_m(t) + \delta\Lambda_m(t_m)$. Define the following notational short hands:

$$\begin{aligned}
M_\delta &:= M_\delta(t, t_m) \text{ or } M_\delta(z, t_m), \\
\Lambda_p &:= \Lambda_p(t) \text{ or } \Lambda_p(z), \\
\Lambda_m &:= \Lambda_m(t_m), \\
\mathcal{L}_m &:= \mathcal{L}_m(M_\delta), \text{ and} \\
\mathcal{L}_p &:= \mathcal{L}_p(\Lambda_p),
\end{aligned}$$

where the argument is z (instead of t) if and only if the function is part of the integrand, and integration is with respect to dz. We also use the same short hands for derivatives of $\mathcal{L}_m$ and $\mathcal{L}_p$, in obvious notation.

We can express $\theta_p(t|t_m)$ in terms of the probability function Q_p defined in Appendix 1.1,

$$\begin{aligned}\theta_p(t|t_m) &= -\frac{\partial}{\partial t}\log\left[-\frac{\partial Q_p(t_m,t)}{\partial t_m}\right]\\ &= \frac{\partial^2 Q_p(t_m,t)/\partial t_m \partial t}{-\partial Q_p(t_m,t)/\partial t_m},\end{aligned} \tag{14}$$

for $t < t_m$. An explicit expression for the numerator of this is easily found by taking the derivative of equation (13) with respect to t_m, which gives (under Case I),

$$\frac{\partial^2 Q_p(t_m,t)}{\partial t_m \partial t} = \delta\lambda_m(t_m)\lambda_p(t)\mathcal{L}'_m[M_\delta(t,t_m)]\mathcal{L}'_p[\Lambda_p(t)], \tag{15}$$

for $t < t_m$. For $t \geq t_m$ the joint density of $T := T_p$ and T_m at (t,t_m) is given by

$$\lambda_m(t_m)\lambda_p(t)\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}'_p[\Lambda_p(t)],$$

so the denominator of (14) is given by

$$\begin{aligned}-\frac{\partial Q_p(t_m,t)}{\partial t_m} &= \int_t^\infty \frac{\partial^2 Q_p(t_m,z)}{\partial t_m \partial z}dz\\ &= \int_t^{t_m} \delta\lambda_m(t_m)\lambda_p(z)\mathcal{L}'_m[M_\delta(z,t_m)]\mathcal{L}'_p[\Lambda_p(z)]dz\\ &\quad + \int_{t_m}^\infty \lambda_m(t_m)\lambda_p(z)\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}'_p[\Lambda_p(z)]dz\\ &= \int_t^{t_m} \delta\lambda_m(t_m)\lambda_p(z)\mathcal{L}'_m[M_\delta(z,t_m)]\mathcal{L}'_p[\Lambda_p(z)]dz\\ &\quad - \lambda_m(t_m)\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)],\end{aligned} \tag{16}$$

for $t < t_m$. Here, we have used that $\lim_{t\to\infty}\Lambda_p(t) = \infty$. Dropping $\lambda_m(t_m)$ from both (15) and (16) gives

$$\theta_p(t|t_m) = \frac{\delta\lambda_p(t)\mathcal{L}'_m\mathcal{L}'_p}{\int_t^{t_m}\delta\lambda_p(z)\mathcal{L}'_m\mathcal{L}'_p dz - \mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)]}, \tag{17}$$

for $t < t_m$. Note that the denominator in (17) is positive due to the fact that derivatives of Laplace transforms are negative.

Expressions for $\partial \log\theta_p(t|t_m)/\partial t$.

First note that $\partial M_\delta(t,t_m)/\partial t = (1-\delta)\lambda_m(t)$. The derivative of the log numerator in equation (17) is

$$\begin{aligned}&\frac{\partial}{\partial t}\log\left[\lambda_m(t_m)^{-1}\frac{\partial^2 Q_p(t_m,t)}{\partial t_m\partial t}\right] =\\ &\quad\frac{\lambda'_p(t)}{\lambda_p(t)} + (1-\delta)\lambda_m(t)\frac{\mathcal{L}''_m[M_\delta(t,t_m)]}{\mathcal{L}'_m[M_\delta(t,t_m)]} + \lambda_p(t)\frac{\mathcal{L}''_p[\Lambda_p(t)]}{\mathcal{L}'_m[\Lambda_p(t)]}.\end{aligned} \tag{18}$$

The derivative of the log denominator in equation (17) is

$$\frac{\partial}{\partial t}\log\left[-\lambda_m(t_m)^{-1}\frac{\partial Q_p(t_m,t)}{\partial t_m}\right] = -\theta_p(t|t_m). \tag{19}$$

Subtracting (19) from (18) gives

$$\frac{\partial \log \theta_p(t|t_m)}{\partial t} = \frac{\lambda_p'(t)}{\lambda_p(t)} + (\delta - 1)\lambda_m(t)\frac{\mathcal{L}_m''}{-\mathcal{L}_m'} - \lambda_p(t)\frac{\mathcal{L}_p''}{-\mathcal{L}_p'} + \theta_p(t|t_m), \tag{20}$$

for $t < t_m$. The second term is positive (negative) if $\delta > 1$ ($\delta < 1$). The first term on the r.h.s. represents individual duration dependence in θ_p, and could be either positive or negative, independent of δ. Also, the combined effect of the last two terms does not necessarily have the sign of $\delta - 1$.

Using $\lim_{t\uparrow t_m} M_\delta(t, t_m) = \Lambda_m(t_m)$ and $\lim_{t\uparrow t_m} \theta_p(t|t_m) = -\delta\lambda_p(t_m)\cdot \mathcal{L}_p'[\Lambda_p(t_m)]/\mathcal{L}_p[\Lambda_p(t_m)]$ gives

$$\begin{aligned}\lim_{t\uparrow t_m} \frac{\partial \log \theta_p(t|t_m)}{\partial t} = &\\ &\frac{\lambda_p'(t_m)}{\lambda_p(t_m)} - \delta\lambda_p(t_m)\left\{\frac{\mathcal{L}_p''[\Lambda_p(t_m)]}{-\mathcal{L}_p'[\Lambda_p(t_m)]} - \frac{-\mathcal{L}_p'[\Lambda_p(t_m)]}{\mathcal{L}_p[\Lambda_p(t_m)]}\right\} \\ &+ (\delta - 1)\left\{\lambda_m(t_m)\frac{\mathcal{L}_m''[\Lambda_m(t_m)]}{-\mathcal{L}_m'[\Lambda_m(t_m)]} + \lambda_p(t_m)\frac{\mathcal{L}_p''[\Lambda_p(t_m)]}{-\mathcal{L}_p'[\Lambda_p(t_m)]}\right\},\end{aligned} \tag{21}$$

but the ambiguity does not disappear. The third term in the right hand side of (21) has the same sign as $\delta - 1$. However, the first term shows that any increase (decrease) in $\theta_p(t|t_m)$ near t_m could be due to positive (negative) duration dependence of $\theta_p(t)$. Furthermore, the second term is (strictly) negative if G_p is not degenerate. This is the observed duration dependence caused by heterogeneity: the expected value of V_p conditional on survival up to time t is decreasing in t.

Let $0 < t_m' < t_m$. Then it follows from (20) and (21) that

$$\begin{aligned}\lim_{t\uparrow t_m'} \left[\frac{\partial \log \theta_p(t|t_m)}{\partial t} - \frac{\partial \log \theta_p(t|t_m')}{\partial t}\right] = &\\ (1 - \delta)\lambda_m(t_m')\left\{\frac{\mathcal{L}_m''[\Lambda_m(t_m')]}{-\mathcal{L}_m'[\Lambda_m(t_m')]} - \frac{\mathcal{L}_m''[M_\delta(t_m', t_m)]}{-\mathcal{L}_m'[M_\delta(t_m', t_m)]}\right\}& \\ + \theta_p(t_m'|t_m) - \lim_{t\uparrow t_m'} \theta_p(t|t_m'),&\end{aligned} \tag{22}$$

for $0 < t_m' < t_m$. Note that $\partial M_\delta(t, t_m)/\partial t_m = \delta\lambda_m(t_m) > 0$ implies that $M_\delta(t, t_m) > M_\delta(t, t) = \Lambda_m(t)$ for all $\delta > 0$ and $0 < t < t_m$. Also, note that $\partial[\mathcal{L}''(s)/\mathcal{L}'(s)]/\partial s = \mathcal{L}'''(s)/\mathcal{L}'(s) - [\mathcal{L}''(s)/\mathcal{L}'(s)]^2 \geq 0$ if $\mathcal{L}$ is a Laplace transform. Using these results, it can easily been seen that the first term on the r.h.s. of (22) has the sign of $1 - \delta$. Next, let

$$\begin{aligned}D_\delta(t_m', t_m) := &- \mathcal{L}_m'[M_\delta(t_m', t_m)]\mathcal{L}_p[\Lambda_p(t_m')] \\ &- \int_{t_m'}^{t_m} \delta\lambda_p(z)\mathcal{L}_m'\mathcal{L}_p' dz + \mathcal{L}_m'[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)].\end{aligned} \tag{23}$$

and note that

$$\lim_{t\uparrow t_m'} \theta_p(t|t_m') = \delta\lambda_p(t_m')\frac{-\mathcal{L}_p'[\Lambda_p(t_m')]}{\mathcal{L}_p[\Lambda_p(t_m')]}. \tag{24}$$

Then, the remaining terms in the r.h.s. of (22) satisfy

$$\theta_p(t'_m|t_m) - \lim_{t \uparrow t'_m} \theta_p(t|t'_m) = \delta\lambda_p(t'_m)\frac{-\mathcal{L}'_p[\Lambda_p(t'_m)]}{\mathcal{L}_p[\Lambda_p(t'_m)]}\times \frac{D_\delta(t'_m, t_m)}{\int_{t'_m}^{t_m} \delta\lambda_p(z)\mathcal{L}'_m\mathcal{L}'_p dz - \mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)]}.$$

We now show that D_δ has the sign of $1-\delta$.

Lemma 1. *Under Case I,*

$$D_\delta(t'_m, t_m) \quad \begin{cases} > 0 \text{ if } \delta < 1 \\ = 0 \text{ if } \delta = 1 \\ < 0 \text{ if } \delta > 1 \end{cases}$$

for all $0 < t'_m < t_m$.

Proof. The integral in D_δ can be expanded by partial integration to

$$\begin{aligned} &\int_{t'_m}^{t_m} \delta\lambda_p(z)\mathcal{L}'_m[M_\delta(z, t_m)]\mathcal{L}'_p[\Lambda_p(z)]dz \\ &\quad = \delta\Big[\mathcal{L}'_m[M_\delta(z, t_m)]\mathcal{L}_p[\Lambda_p(z)]\Big]_{t'_m}^{t_m} \\ &\qquad - \int_{t'_m}^{t_m} \delta(1-\delta)\lambda_m(z)\mathcal{L}''_m[M_\delta(z, t_m)]\mathcal{L}_p[\Lambda_p(z)]dz \\ &\quad = \delta\Big\{\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)] - \mathcal{L}'_m[M_\delta(t'_m, t_m)]\mathcal{L}_p[\Lambda_p(t'_m)]\Big\} \\ &\qquad - \int_{t'_m}^{t_m} \delta(1-\delta)\lambda_m(z)\mathcal{L}''_m[M_\delta(z, t_m)]\mathcal{L}_p[\Lambda_p(z)]dz. \end{aligned} \tag{25}$$

Substituting (25) in (23) gives

$$\begin{aligned} D_\delta(t'_m, t_m) = &(1-\delta)\Big\{-\mathcal{L}'_m[M_\delta(t'_m, t_m)]\mathcal{L}_p[\Lambda_p(t'_m)] + \mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)] \\ &+ \int_{t'_m}^{t_m} \delta\lambda_m(z)\mathcal{L}''_m[M_\delta(z, t_m)]\mathcal{L}_p[\Lambda_p(z)]dz.\Big\} \end{aligned} \tag{26}$$

Note that $\partial M_\delta(t, t_m)/\partial t = (1-\delta)\lambda_m(t)$ implies that $M_\delta(t'_m, t_m) < M_\delta(t_m, t_m) = \Lambda_m(t_m)$, and therefore that the term between brackets in (26) is positive, for all $\delta < 1$ and $0 < t'_m < t_m$. So, for $\delta \leq 1$ it is immediately clear that D_δ has the sign of $1-\delta$, and it only remains to be shown that the term between brackets is also positive for $\delta > 1$. If $\delta > 1$, we can bound this term from below by

$$
\begin{aligned}
&-\mathcal{L}'_m[M_\delta(t'_m,t_m)]\mathcal{L}_p[\Lambda_p(t'_m)]+\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)] \\
&\qquad\qquad\qquad\qquad + \int_{t'_m}^{t_m} \delta\lambda_m(z)\mathcal{L}''_m[M_\delta(z,t_m)]\mathcal{L}_p[\Lambda_p(z)]dz \\
&\geq -\mathcal{L}'_m[M_\delta(t'_m,t_m)]\mathcal{L}_p[\Lambda_p(t_m)]+\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)] \\
&\qquad\qquad\qquad\qquad + \int_{t'_m}^{t_m} \delta\lambda_m(z)\mathcal{L}''_m[M_\delta(z,t_m)]\mathcal{L}_p[\Lambda_p(t_m)]dz \\
&= -\mathcal{L}'_m[M_\delta(t'_m,t_m)]\mathcal{L}_p[\Lambda_p(t_m)]+\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)] \\
&\qquad\qquad\qquad\qquad - \frac{\delta}{\delta-1}\mathcal{L}_p[\Lambda_p(t_m)]\left\{\mathcal{L}'_m[\Lambda_m(t_m)]-\mathcal{L}'_m[M_\delta(t'_m,t_m)]\right\} \\
&= \frac{1}{\delta-1}\mathcal{L}_p[\Lambda_p(t_m)]\left\{-\mathcal{L}'_m[\Lambda_m(t_m)]+\mathcal{L}'_m[M_\delta(t'_m,t_m)]\right\} \geq 0,
\end{aligned}
$$

where the last inequality follows from the fact that $M_\delta(t'_m,t_m) > M_\delta(t_m,t_m) = \Lambda_m(t_m)$ if $\delta > 1$. □

As a result, $\theta_p(t'_m|t_m) - \lim_{t\uparrow t'_m}\theta_p(t|t'_m)$ also has the sign of $1-\delta$, and the whole expression in (22) has the sign of $1-\delta$. In sum, if $\delta > 1$ ($\delta < 1$) then $\log\theta_p(t|t'_m)$ increases (decreases) relative to $\log\theta_p(t|t_m)$, for t near t'_m and all $0 < t'_m < t_m$.

Now let us turn to the effect of t_m on the level of $\theta_p(t|t_m)$. Lemma 1 is also of direct use here. Evaluating (17) at $t = t'_m$, dividing by (24), and taking logs gives

$$
\begin{aligned}
\lim_{t\uparrow t'_m}\log\left[\frac{\theta_p(t|t_m)}{\theta_p(t|t'_m)}\right] &= -\log\left\{\int_{t'_m}^{t_m}\delta\lambda_p(z)\mathcal{L}'_m\mathcal{L}'_p dz - \mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)]\right\} \\
&\quad + \log\left\{-\mathcal{L}'_m[M_\delta(t'_m,t_m)]\mathcal{L}_p[\Lambda_p(t'_m)]\right\}.
\end{aligned}
\tag{27}
$$

Clearly, $\lim_{t\uparrow t'_m}\log[\theta_p(t|t_m)/\theta_p(t|t'_m)]$ has the same sign as $D_\delta(t'_m,t_m)$ and thus, by Lemma 1, as $1-\delta$.

This proves Proposition 1.

Appendix 1.3 Additional results for Case I

Expressions for $\partial\log\theta_p(t|t_m)/\partial t_m$.

Note that $\partial M_\delta(t,t_m)/\partial t_m = \delta\lambda_m(t_m)$, $M_\delta(t,t) = \Lambda_m(t)$, and

$$
\begin{aligned}
&\frac{\partial}{\partial t_m}\left[-\lambda_m(t_m)^{-1}\frac{\partial Q_p(t_m,t)}{\partial t_m}\right] = \\
&\qquad (\delta-1)\lambda_p(t_m)\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}'_p[\Lambda_p(t_m)] \\
&\qquad + \int_t^{t_m}\delta^2\lambda_p(z)\lambda_m(t_m)\mathcal{L}''_m[M_\delta(z,t_m)]\mathcal{L}'_p[\Lambda_p(z)]dz \\
&\qquad - \lambda_m(t_m)\mathcal{L}''_m[\Lambda_m(t_m)]\mathcal{L}_p[\Lambda_p(t_m)].
\end{aligned}
\tag{28}
$$

The derivative of the log denominator in (17) is given by

$$
\frac{\partial}{\partial t_m}\log\left[\lambda_m(t_m)^{-1}\frac{\partial^2 Q_p(t_m,t)}{\partial t_m\partial t}\right] = \delta\lambda_m(t_m)\frac{\mathcal{L}''_m[M_\delta(t,t_m)]}{\mathcal{L}'_m[M_\delta(t,t_m)]}
\tag{29}
$$

The derivative of (17) with respect to t_m is then given by

$$\frac{\partial \log \theta_p(t|t_m)}{\partial t_m} = \frac{\partial}{\partial t_m} \log \left[\lambda_m(t_m)^{-1} \frac{\partial^2 Q_p(t_m,t)}{\partial t_m \partial t} \right] - \frac{\frac{\partial}{\partial t_m}\left[-\lambda_m(t_m)^{-1} \frac{\partial Q_p(t_m,t)}{\partial t_m} \right]}{\left[-\lambda_m(t_m)^{-1} \frac{\partial Q_p(t_m,t)}{\partial t_m} \right]} \tag{30}$$

To evaluate the sign of this derivative, we multiply (30) with the positive numerator $\left[-\lambda_m(t_m)^{-1}\partial Q_p(t_m,t)/\partial t_m\right]$, which gives

$$\begin{aligned}
&\left[-\lambda_m(t_m)^{-1} \frac{\partial Q_p(t_m,t)}{\partial t_m} \right] \frac{\partial \log \theta_p(t|t_m)}{\partial t_m} = \\
&\quad (1-\delta)\lambda_p(t_m)\mathcal{L}'_m[\Lambda_m(t_m)]\mathcal{L}'_p[\Lambda_p(t_m)] + \delta^2 \lambda_m(t_m) \times \\
&\quad \int_t^{t_m} \lambda_p(z)\mathcal{L}'_p[\Lambda_p(z)]\mathcal{L}'_m[M_\delta(z,t_m)] \left\{ \frac{\mathcal{L}''_m[M_\delta(t,t_m)]}{\mathcal{L}'_m[M_\delta(t,t_m)]} - \frac{\mathcal{L}''_m[M_\delta(z,t_m)]}{\mathcal{L}'_m[M_\delta(z,t_m)]} \right\} dz \\
&\quad - \lambda_m(t_m)\mathcal{L}_p[\Lambda_p(t_m)]\mathcal{L}'_m[\Lambda_m(t_m)] \left\{ \delta \frac{\mathcal{L}''_m[M_\delta(t,t_m)]}{\mathcal{L}'_m[M_\delta(t,t_m)]} - \frac{\mathcal{L}''_m[\Lambda_m(t_m)]}{\mathcal{L}'_m[\Lambda_m(t_m)]} \right\}.
\end{aligned} \tag{31}$$

Recall that $\partial M_\delta(t,t_m)/\partial t = (1-\delta)\lambda_m(t)$ implies that $M_\delta(t,t_m) > M_\delta(z,t_m)$ if $\delta > 1$ and $M_\delta(t,t_m) < M_\delta(z,t_m)$ if $\delta < 1$, for $t < z \leq t_m$. In particular, $M_\delta(t,t_m) > M_\delta(t_m,t_m) = \Lambda_m(t_m)$ if $\delta > 1$ and $M_\delta(t,t_m) < \Lambda_m(t_m)$ if $\delta < 1$, for $t < t_m$. Also, note that $\partial[\mathcal{L}''(s)/\mathcal{L}'(s)]/\partial s = \mathcal{L}'''(s)/\mathcal{L}'(s) - [\mathcal{L}''(s)/\mathcal{L}'(s)]^2 \geq 0$ if $\mathcal{L}$ is a Laplace transform. Collecting these results, it is clear that the term between brackets in the integrand of the second term on the r.h.s. has the same sign as $\delta - 1$. Again, this causes ambiguity, as the first term in the r.h.s. has the sign of $1-\delta$.

Expressions for $\partial^2 \log \theta_p(t|t_m)/\partial t \partial t_m$.

The cross-derivative of $\log \theta_p(t|t_m)$ is given by

$$\begin{aligned}
\frac{\partial^2 \log \theta_p(t|t_m)}{\partial t \partial t_m} =& \delta(1-\delta)\lambda_m(t)\lambda_m(t_m) \left\{ \frac{\mathcal{L}'''_m[M_\delta(t,t_m]}{\mathcal{L}'_m[M_\delta(t,t_m]} - \left(\frac{\mathcal{L}''_m[M_\delta(t,t_m]}{\mathcal{L}'_m[M_\delta(t,t_m]} \right)^2 \right\} \\
&+ \frac{\partial \theta_p(t|t_m)}{\partial t_m}.
\end{aligned} \tag{32}$$

The first term in the r.h.s. of (32) has the sign of $1-\delta$, but the second term has the sign of $\partial \log \theta_p(t|t_m)/\partial t_m$, which we cannot sign unambiguously, as argued before.

An example.

Let $\delta = 1/2$, $\lambda_m(t) \equiv 1$ and $\lambda_p(t) \equiv 1$. Furthermore, let G_m be discrete with support $\{1,2\}$ and $\Pr(V_m = 1) = \Pr(V_m = 2) = 1/2$, and let G_p be degenerate at 1. Then, $\mathcal{L}_m(s) = [\exp(-s) + \exp(-2s)]/2$, $\mathcal{L}_p(s) = \exp(-s)$, and $\Lambda_m(t) = \Lambda_p(t) = t$. For simplicity we give results for $t \downarrow 0$. It can be shown that the numerator and denominator of $\theta_p(t|t_m)$ satisfy

$$-\lim_{t\downarrow 0} \lambda_m(t_m)^{-1} \frac{\partial Q_p(t_m,t)}{\partial t_m} = \frac{4\exp(-2t_m) + 9\exp(-3t_m) + 2\exp(-t_m/2) + 3\exp(-t_m)}{12}, \tag{33}$$

and

$$\lim_{t\downarrow 0} \lambda_m(t_m)^{-1} \frac{\partial^2 Q_p(t_m,t)}{\partial t_m \partial t} = \frac{\exp(-t_m/2) + 2\exp(-t_m)}{4}, \tag{34}$$

respectively. Dividing (34) by (33) gives

$$\theta_p(0+|t_m) = \frac{3\exp(5t_m/2) + 6\exp(2t_m)}{4\exp(t_m) + 9 + 2\exp(5t_m/2) + 3\exp(2t_m)}. \tag{35}$$

Note that $\lim_{t_m \downarrow 0} \theta_p(0+, t_m) = 1/2$ and $\lim_{t_m \to \infty} \theta_p(0+, t_m) = 3/2$. As a function of t_m, $\lim_{t \downarrow 0} \theta_p(t|t_m)$ is increasing for small t_m and decreasing for large t_m. This can also be learned directly from the fact that, for $t \downarrow 0$, the expression in (31) equals

$$\left\{[24\exp(t_m/2) + 48]\,[\exp(-t_m/2) + 2\exp(-t_m)]\right\}^{-1} \times \Big[12\exp(-2t_m) + 40\exp(-5t_m/2)+$$
$$77\exp(-3t_m) + 162\exp(-7t_m/2) + 144\exp(-4t_m) - \exp(-t_m) - 2\exp(-3t_m/2)\Big]$$

in this case. Multiplying with (33) gives

$$\frac{\partial \log \theta_p(0+|t_m)}{\partial t_m} = \frac{12\exp(t_m) + 40\exp(t_m/2) + 77 + 162\exp(-t_m/2) + 144\exp(-t_m) - \exp(2t_m) - 2\exp(3t_m/2)}{[2\exp(t_m/2) + 4]\,[\exp(-t_m/2) + 2\exp(-t_m)]\,[4\exp(t_m) + 9 + 2\exp(5t_m/2) + 3\exp(2t_m)]}. \tag{36}$$

Now note that $\lim_{t_m \downarrow 0} \theta_p(0+, t_m) = 4/3$ and $\lim_{t_m \to \infty} \exp(t_m/2)\theta_p(0+, t_m) = -1/4$. So, $\theta_p(0+, t_m)$ is positive near 0 and negative for sufficiently large t_m. More precisely, $\exp(t_m/2)\theta_p(0+, t_m)$ is monotonically decreasing from $4/3$ at 0 to $-1/4$ at infinity. So, we can conclude that we get ambiguous results on the sign of $\partial \log \theta_p(t|t_m)/\partial t_m$ for fixed $\delta < 1$.

It is easy to extend this example to $\partial^2 \log \theta_p(t|t_m)/\partial t \partial t_m$ by noting that the second term in the r.h.s. of (32) equals $\theta_p(t|t_m)\partial \log[\theta_p(t|t_m)]/\partial t_m$. As $\theta_p(0+, t_m)$ is positive and has a finite limit as $t_m \to \infty$, for $t \downarrow 0$ this term behaves much like $\partial \log[\theta_p(0+|t_m)]/\partial t_m$. The first term on the r.h.s. of (32) has the sign of $1 - \delta$, but can be suppressed for $t \downarrow 0$ by changing λ_p in a neighborhood of 0 such that $\lambda_p(0+) = 0$. The effects of this change on the results for the second term can be made arbitrarily small by changing λ_p on a sufficiently small neighborhood of 0. Concluding, the first term can be made to vanish without changing much of the results concerning the second term, for $t \downarrow 0$, and the ambiguity result concerning the first derivative carries over to the cross-derivative. A specific example that produces the desired ambiguity in both derivatives is the example above with $\lambda_p(t) = 1/10$ for $t < 1$ and $\lambda_p(t) = 1$ for $t \geq 1$.

Appendix 1.4 Proof of Proposition 2

If G_m is degenerate, say at 1, then $\mathcal{L}_m^{(i)}(s) = (-1)^i \exp(-s)$. This implies that $\mathcal{L}''_m(s)/\mathcal{L}'_m(s) = -1$, and is independent of s. Therefore, the second term in the r.h.s. of (31) vanishes, and the third term reduces to $-(1-\delta)\mathcal{L}_p[\Lambda_p(t_m)]\mathcal{L}'_m[\Lambda_m(t_m)]$, which carries the sign of $1-\delta$. Therefore, $\partial \log \theta_p(t|t_m)/\partial t_m$ has the sign of $1-\delta$, which proves the first assertion in Proposition 2. The proof of the second assertion easily follows from the observation that $\mathcal{L}'''_m(s)/\mathcal{L}'_m(s) - [\mathcal{L}''_m(s)/\mathcal{L}'_m(s)]^2 = 0$ for all s, as G_m is degenerate. Substitution in equation (32) shows that $\partial^2 \log \theta_p(t|t_m)/\partial t \partial t_m = \partial \theta_p(t|t_m)/\partial t_m$, which has the same sign as $\partial \log \theta_p(t|t_m)/\partial t_m$.

Appendix 2 Proofs for Case II: pure selection

In this appendix we analyze a generalized version of Case II, denoted by Case IIa (recall that Case Ia is actually a more *restrictive* version of Case I).

Appendix 2.1 Definition of the more general Case IIa

Case IIa: $\delta \equiv 1$. Furthermore, the joint distribution G of (V_m, V_p) is such that $V_p = h(V_m)$ for some nonnegative h which is regularly varying at 0 with exponent $\tau \geq 0$,

$$h(v) \stackrel{v\downarrow 0}{\sim} v^{\tau} M(v),$$

where M is an arbitrary function that is slowly varying at 0. The random variable V_m has a distribution function F concentrated on $(0, \infty)$. We denote F_t to be the class of (improper) distribution functions that satisfy $dF_t(v) := \exp[-\Lambda_p(t)h(v)]dF(v)$. We assume that the model determinants are such that F_t satisfies either one of the following conditions.

(i). F_t is absolutely continuous with density f_t. Let $t \geq 0$ be given. Then, f_t is regularly varying at 0 with exponent $-1 < \sigma_t < \infty$, or

$$f_t(v) \stackrel{v\downarrow 0}{\sim} v^{\sigma_t} L_t(v),$$

where L_t is slowly varying at 0.

(ii). F_t is infinitely discrete, i.e. concentrated on a countable subset $S \subset (0, \infty)$, with dense support near 0. Denote the elements of S by the sequence $v_0 > v_1 > v_2 > \cdots > 0$, and let[17]

$$dF_t(v) = \begin{cases} p_{k|t} & \text{if } v = v_k, \; k = 0, 1, 2, \cdots \\ 0 & \text{elsewhere} \end{cases}$$

for some $0 < p_{k|t} < 1$, $k = 0, 1, 2, \cdots$. Let $t \geq 0$ be given. Then, p_k is regularly varying at infinity with exponent $-\infty < -\sigma_t < -1$, or

$$p_{k|t} \stackrel{k\to\infty}{\sim} k^{-\sigma_t} L_t(k),$$

where L_t is slowly varying at ∞. Furthermore, v_k is regularly varying at infinity with exponent $-\infty < -\sigma < 0$, or

$$v_k \stackrel{k\to\infty}{\sim} k^{-\sigma} L(k),$$

where L is slowly varying at infinity.

(which is where the definition of Case IIa ends.) In the proofs, we focus on continuous distributions of type (i). Discrete distributions of type (ii) can basically be treated in the same way.

[17]Note that we are focusing on the behavior of F_t near 0, and that the restriction to a support bounded from above by v_0 is immaterial and only made for notational convenience.

Appendix 2.2 Proof for Case IIa and Proof of Proposition 3

It can easily be shown that, in the notation introduced in Appendix 1.1,

$$\theta_p(t|t_m) = -\lambda_p(t)\frac{\mathcal{L}_G^{(m,p)}[\Lambda_m(t_m),\Lambda_p(t)]}{\mathcal{L}_G^{(m)}[\Lambda_m(t_m),\Lambda_p(t)]}, \tag{37}$$

where

$$\mathcal{L}_G^{(m^i,p^j)}(s,t) := \frac{\partial^{i+j}\mathcal{L}_G(s,t)}{\partial s^i \partial t^j}$$

denotes the (i,j)-th cross-derivative of $\mathcal{L}_G$. Note that these exist for all $t, t_m > 0$. Taking the derivatives to t_m and then to t of (37) in logs gives

$$\frac{\partial \log\theta_p(t|t_m)}{\partial t_m} = \lambda_m(t_m)\left[\frac{\mathcal{L}_G^{(m^2,p)}[\Lambda_m(t_m),\Lambda_p(t)]}{\mathcal{L}_G^{(m,p)}[\Lambda_m(t_m),\Lambda_p(t)]} - \frac{\mathcal{L}_G^{(m^2)}[\Lambda_m(t_m),\Lambda_p(t)]}{\mathcal{L}_G^{(m)}[\Lambda_m(t_m),\Lambda_p(t)]}\right] \tag{38}$$

and

$$\begin{aligned}\frac{\partial^2 \log\theta_p(t|t_m)}{\partial t\partial t_m} &= \lambda_p(t)\lambda_m(t_m)\times \\ &\left[\frac{\mathcal{L}_G^{(m^2)}(\cdot)\mathcal{L}_G^{(m,p)}(\cdot)}{\mathcal{L}_G^{(m)}(\cdot)^2} - \frac{\mathcal{L}_G^{(m^2,p)}(\cdot)}{\mathcal{L}_G^{(m)}(\cdot)}\right. \\ &\left. + \frac{\mathcal{L}_G^{(m^2,p^2)}(\cdot)}{\mathcal{L}_G^{(m,p)}(\cdot)} - \frac{\mathcal{L}_G^{(m^2,p)}(\cdot)\mathcal{L}_G^{(m,p^2)}(\cdot)}{\mathcal{L}_G^{(m,p)}(\cdot)^2}\right].\end{aligned} \tag{39}$$

It will prove useful to rewrite equations (37), (38) and (39) in terms of the conditional moments $\mu_{i,j} := \mathrm{E}[V_m^i V_p^j | T_m \geq t_m, T_p \geq t]$, which are given by

$$\mu_{i,j}(t_m,t) = (-1)^{i+j}\frac{\mathcal{L}_G^{(m^i,p^j)}[\Lambda_m(t_m),\Lambda_p(t)]}{\mathcal{L}_G[\Lambda_m(t_m),\Lambda_p(t)]}.$$

Also, using the normalized moments

$$\gamma_{i,j}(t_m,t) := \frac{\mu_{i,j}(t_m,t)}{\mu_{1,0}(t_m,t)^i\mu_{0,1}(t_m,t)^j}$$

we can write

$$\theta_p(t|t_m) = \lambda_p(t)\mu_{0,1}(t_m,t)\gamma_{1,1}(t_m,t), \tag{40}$$

$$\frac{\partial \log\theta_p(t|t_m)}{\partial t_m} = \lambda_m(t_m)\mu_{1,0}(t_m,t)\left[\gamma_{2,0}(t_m,t) - \frac{\gamma_{2,1}(t_m,t)}{\gamma_{1,1}(t_m,t)}\right] \tag{41}$$

and

$$\begin{aligned}\frac{\partial^2 \log\theta_p(t|t_m)}{\partial t\partial t_m} &= \lambda_p(t)\lambda_m(t_m)\mu_{1,0}(t_m,t)\mu_{0,1}(t_m,t)\times \\ &\left[\gamma_{2,0}(t_m,t)\gamma_{1,1}(t_m,t) - \gamma_{2,1}(t_m,t) + \frac{\gamma_{2,2}(t_m,t)}{\gamma_{1,1}(t_m,t)} - \frac{\gamma_{2,1}(t_m,t)\gamma_{1,2}(t_m,t)}{\gamma_{1,1}(t_m,t)^2}\right].\end{aligned} \tag{42}$$

We now present some lemmas that are used in proving the results for the absolutely continuous version of Case IIa. First, we need two results from the theory of regular variation. The first lemma mirrors part of the lemma from Feller (1971), Section VIII.9.

Lemma 2. *Let $L > 0$ vary slowly at 0. Then, for $i > -1$, $\int_0^v \eta^i L(\eta) d\eta$ varies regularly at 0 with exponent $i + 1$.*

Proof. Define $F_i(v) := \int_0^v \eta^i L(\eta) d\eta$. As L is slowly varying, for each $\alpha > 0$ and $\epsilon > 0$, $(1 - \epsilon)L(v) < L(\alpha v) < (1 + \epsilon)L(v)$ for all sufficiently small v. Therefore,

$$(1 - \epsilon)\alpha^{i+1} F_i(v) \leq F_i(\alpha v) \leq (1 + \epsilon)\alpha^{i+1} F_i(v) \tag{43}$$

for all sufficiently small v. Note that from (43) it is clear that F_i converges for $v \downarrow 0$ for all $i > -1$. Dividing (43) by $F_i(v)$ gives

$$(1 - \epsilon)\alpha^{i+1} \leq \frac{F_i(\alpha v)}{F_i(v)} \leq (1 + \epsilon)\alpha^{i+1}.$$

Because this is true for arbitrary $\epsilon > 0$ the claimed result follows. □

Next, let F be a distribution function and let F_t denote a class of (improper) distribution functions such that $dF_t(v) := \exp[-\Lambda_p(t)h(v)]dF(v)$. Let $F_{i,j}(v|t) := \int_0^v \eta^i h(\eta)^j dF_t(\eta)$, and assume that F_t is absolutely continuous with density f_t. By analogy to Feller (1971), Section VIII.9, Theorem 1,

Lemma 3. *Let $t > 0$ be given. If $L_t > 0$ and $M > 0$ are slowly varying at 0, $-1 < \sigma_t < \infty$, $0 \leq \tau < \infty$, $f_t(v) \overset{v\downarrow 0}{\sim} v^{\sigma_t} L_t(v)$, and $h(v) \overset{v\downarrow 0}{\sim} v^{\tau} M(v)$, then*

$$F_{i,j}(v|t) \quad \overset{v\downarrow 0}{\sim} \quad v^{\sigma_t+1+i+\tau j} \frac{L_t(v)M(v)^j}{\sigma_t + 1 + i + \tau j}, \text{ for } \sigma_t + 1 + i + \tau j > 0.$$

Proof. Let $Z(v)$ be such that

$$\frac{Z(v)}{v} = \frac{v^i h(v)^j f_t(v)}{F_{i,j}(v|t)}. \tag{44}$$

As f_t varies regularly with exponent σ_t, $h(v)$ varies regularly with exponent τ, and $\sigma_t + 1 + i + \tau j > 0$, $v^i h(v)^j f_t(v)$ varies regularly with exponent $\sigma_t + i + \tau j$, and, by Lemma 2, $F_{i,j}$ varies regularly with exponent $\sigma_t + 1 + i + \tau j$. Therefore, $Z(v)/v$ varies regularly with exponent -1, and $Z(v)$ is slowly varying at 0. As the numerator in the r.h.s. of (44) is almost everywhere the derivative of the denominator, we can integrate between v and αv, which gives

$$\begin{aligned} \log \frac{F_{i,j}(\alpha v|t)}{F_{i,j}(v|t)} &= \int_v^{\alpha v} \frac{Z(\eta)}{\eta} d\eta \\ &= \int_1^{\alpha} \frac{Z(\eta v)}{\eta} d\eta \\ &= Z(v) \int_1^{\alpha} \frac{Z(\eta v)}{Z(v)} \frac{1}{\eta} d\eta. \end{aligned} \tag{45}$$

Due to regular variation, the l.h.s. of (45) tends to $(\sigma_t + 1 + i + \tau j) \log \alpha$ if $v \downarrow 0$. Let $v \downarrow 0$ such that the integral in the last line converges to a limit $l \leq \infty$. Because Z is slowly varying, the integrand in the last line converges to η^{-1} if $v \downarrow 0$. Therefore, by Fatou's Lemma, $l \geq \log \alpha$. This implies that $Z(v) \to m$ if $v \downarrow 0$, where $m \leq \sigma_t + 1 + i + \tau j < \infty$. Therefore, Z is bounded near 0, and $\lim_{v\downarrow 0} Z(\alpha v) = m$ for all α. As the integrand in the second line is bounded for sufficiently small v, $\lim_{v\downarrow 0} \int_1^{\alpha} [Z(\eta v)/\eta] d\eta = m \log \alpha$, and therefore $m = \sigma_t + 1 + i + \tau j$. Combining with (44) gives the desired result. □

Without proof we state the following well known Abelian theorem (see e.g. Feller, 1971, Section XIII.5, Theorem 3).

Lemma 4. *If L is slowly varying at 0 and $0 \leq \rho < \infty$, then $H(v) \overset{v\downarrow 0}{\sim} v^{\rho} L(v)$ implies $\mathcal{L}_H(s) \overset{s\to\infty}{\sim} \Gamma(\rho+1)s^{-\rho}L(1/s)$.*

We can apply this to $F_{i,j}$.

Lemma 5. *Let $t > 0$ be given. If $L_t > 0$ and $M > 0$ are slowly varying at 0, $-1 < \sigma_t < \infty$, $0 \leq \tau < \infty$, $f_t(v) \overset{v\downarrow 0}{\sim} v^{\sigma_t} L_t(v)$, and $h(v) \overset{v\downarrow 0}{\sim} v^{\tau} M(v)$, then*

$$\mathcal{L}_{F_{i,j}|t}[\Lambda_m(t_m)] \overset{t_m\to\infty}{\sim} \Gamma(\sigma_t+2+i+\tau j)\Lambda_m(t_m)^{-(\sigma_t+1+i+\tau j)} \times \frac{L_t[\Lambda_m(t_m)^{-1}]M[\Lambda_m(t_m)^{-1}]^j}{\sigma_t+1+i+\tau j},$$

where $\mathcal{L}_{F_{i,j}|t}$ is the Laplace transform of $F_{i,j}(v|t)$.

Proof. This is a direct consequence of Lemmas 3 and 4. □

As a result of this lemma, we have

Lemma 6. *Let $t > 0$ be given. If $L_t > 0$ and $M > 0$ are slowly varying at 0, $-1 < \sigma_t < \infty$, $0 \leq \tau < \infty$, $f_t(v) \overset{v\downarrow 0}{\sim} v^{\sigma_t} L_t(v)$, and $h(v) \overset{v\downarrow 0}{\sim} v^{\tau} M(v)$, then*

$$\mu_{i,j}(t_m, t) \overset{t_m\to\infty}{\sim} \frac{\Gamma(\sigma_t+1+i+\tau j)}{\Gamma(\sigma_t+1)}\Lambda_m(t_m)^{-(i+\tau j)}M[\Lambda_m(t_m)^{-1}]^j,$$

and therefore

$$\gamma_{i,j}(\infty, t) := \lim_{t_m\to\infty} \gamma_{i,j}(t_m, t) = \frac{\Gamma(\sigma_t+1+i+\tau j)}{(\sigma+1)^i\Gamma(\sigma_t+1)}\left(\frac{\Gamma(\sigma_t+1)}{\Gamma(\sigma_t+1+\tau)}\right)^j.$$

Proof. Follows directly from Lemma 5. □

We are now in a position to prove

Proposition 4. *In Case IIa, with a non-degenerate F, there holds for each $t > 0$ that*

$$\lim_{t_m\to\infty} \frac{\Lambda_m(t_m)}{\lambda_m(t_m)} \frac{\partial \log\theta_p(t|t_m)}{\partial t_m} = -\tau \leq 0,$$

and

$$\lim_{t_m\to\infty} \frac{\Lambda_m(t_m)^{\tau+1}}{\lambda_m(t_m)M[\Lambda_m(t_m)^{-1}]} \frac{\partial^2 \log\theta_p(t|t_m)}{\partial t\partial t_m} =$$

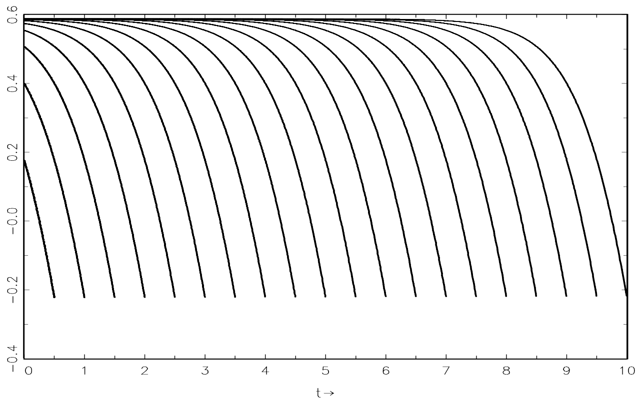
$$\tau\lambda_p(t)\left[\frac{\Gamma(\sigma_t+2+2\tau)}{\Gamma(\sigma_t+2+\tau)} - \frac{\Gamma(\sigma_t+2+\tau)}{\Gamma(\sigma_t+2)}\right] \begin{cases} > 0 \text{ if } \tau > 0 \\ = 0 \text{ if } \tau = 0 \end{cases}$$

Proof. Follows directly from Lemma 6. The last assertion follows from log-convexity of the Gamma function. □

Finally, note that in Case II, $h(v) = v^{\tau}$, so that Case II is the special case of Case IIa with $M(v) = 1$, $\sigma_t = \sigma$ and $L_t(v) = \exp[-\Lambda_p(t)v^{\tau}]L(v)$.[18] It is easy to check that Proposition 4 reduces to Proposition 3 under these restrictions, which proves Proposition 3.

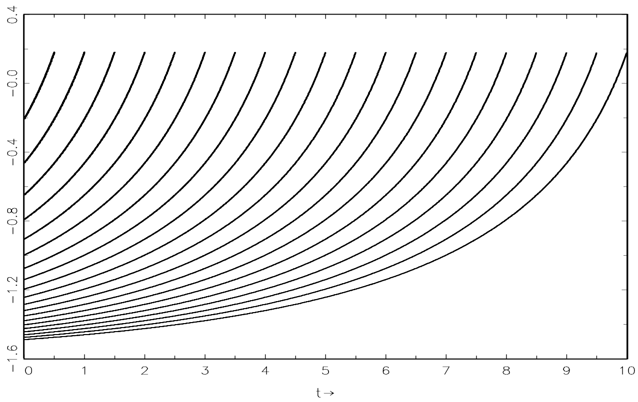
[18] We could also simply take $L_t(v) = L(v)$ if $\tau > 0$, and $L_t(v) = \exp[-\Lambda_p(t)]L(v)$ if $\tau = 0$.

Figure 1: $\log \theta_p(t|t_m)$ for various values of t_m (Case I: $\delta = 0.8$; G_m and G_p degenerate at 1)



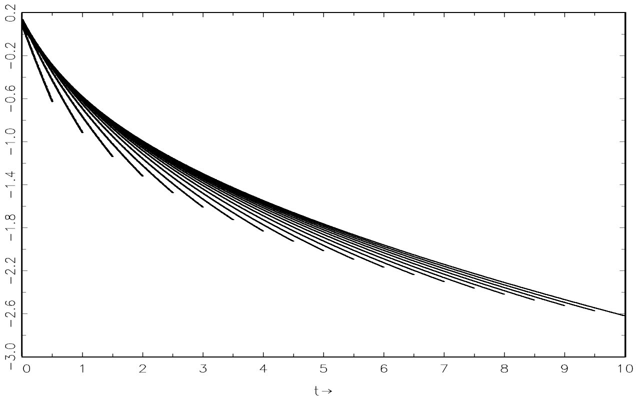
Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

Figure 2: $\log \theta_p(t|t_m)$ for various values of t_m (Case I: $\delta = 1.2$; G_m and G_p degenerate at 1)



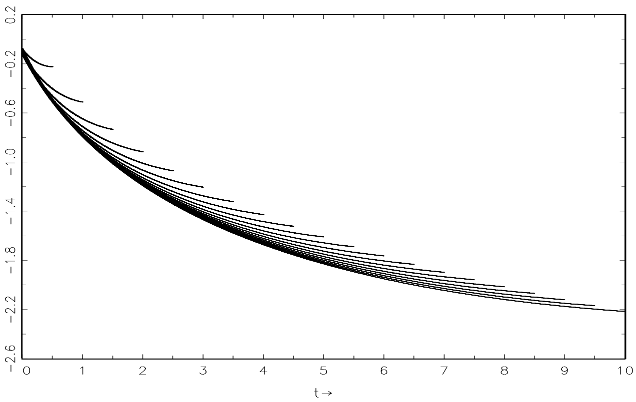
Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

Figure 3: $\log \theta_p(t|t_m)$ for various values of t_m (Case I: $\delta = 0.8$; G_m and G_p unit exponential)



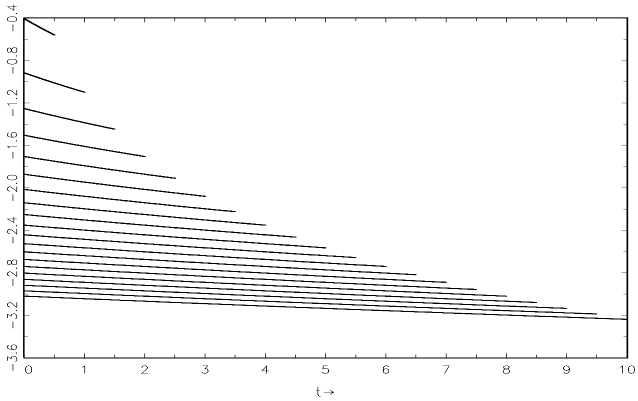
Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

Figure 4: $\log \theta_p(t|t_m)$ for various values of t_m (Case I: $\delta = 1.2$; G_m and G_p unit exponential)



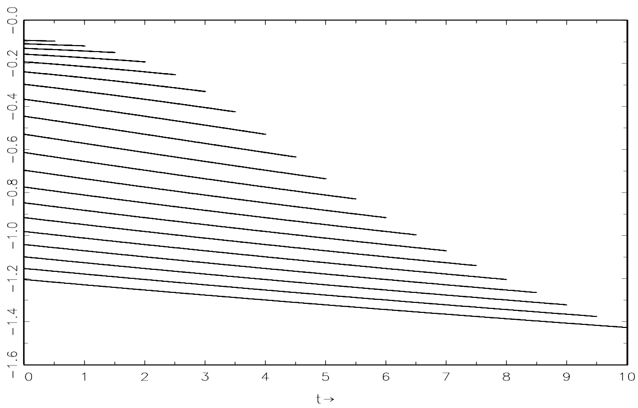
Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

Figure 5: $\log \theta_p(t|t_m)$ for various values of t_m (Case II: $f(v) = \exp(-v)$; $V_p = V_m$)



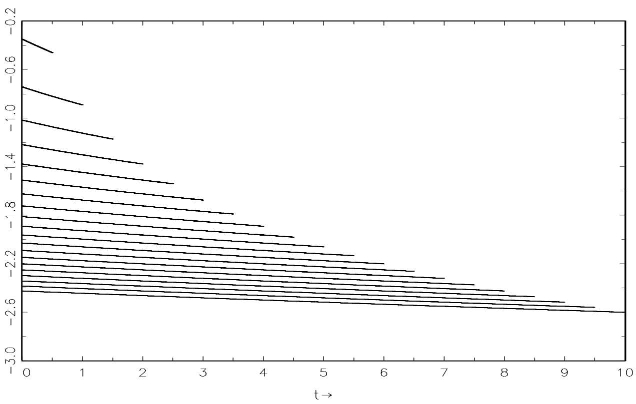
Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

Figure 6: $\log \theta_p(t|t_m)$ for various values of t_m (Case II: $f(v) = 11v^{10}$; $V_p = V_m$)



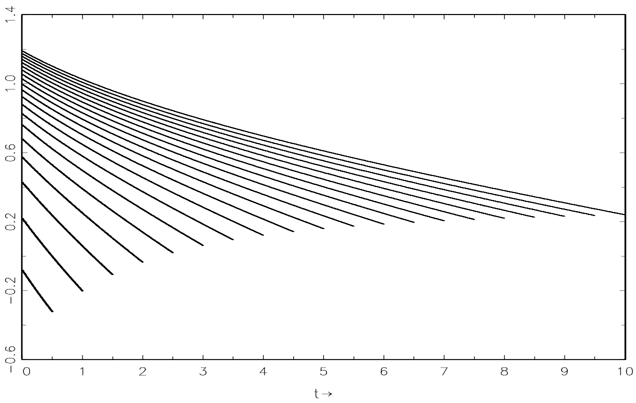
Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

Figure 7: $\log\theta_p(t|t_m)$ for various values of t_m (Case II: $f(v) \propto \exp(-(\log v)^2/2)/v$; $V_p = V_m$; $0 < v < 2$)



Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

Figure 8: $\log\theta_p(t|t_m)$ for various values of t_m (Case II: $f(v) \propto \exp(-v)$; $V_p = -\log V_m$; $0 < v < 1$)



Note: Each curve corresponds to a t_m equal to the largest value of t for which it is drawn.

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
796	P. Apps	Gender, Time Use and Models of the Household	5	06/03
797	E. Bratberg Ø. A. Nilsen K. Vaage	Assessing Changes in Intergenerational Earnings Mobility	2	06/03
798	J. J. Heckman J. A. Smith	The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program	6	06/03
799	R. A. Hart	General Human Capital and Employment Adjustment in the Great Depression: Apprentices and Journeymen in UK Engineering	2	06/03
800	T. Beissinger C. Knoppik	Sind Nominallöhne starr? Neuere Evidenz und wirtschaftspolitische Implikationen	7	06/03
801	A. Launov	A Study of the Austrian Labor Market Dynamics Using a Model of Search Equilibrium	2	06/03
802	H. Antecol P. Kuhn S. J. Trejo	Assimilation via Prices or Quantities? Labor Market Institutions and Immigrant Earnings Growth in Australia, Canada, and the United States	1	06/03
803	R. Lalive	Social Interactions in Unemployment	3	06/03
804	J. H. Abbring	Dynamic Econometric Program Evaluation	6	06/03
805	G. J. van den Berg A. van Vuuren	The Effect of Search Frictions on Wages	6	06/03
806	G. J. van den Berg	Multiple Equilibria and Minimum Wages in Labor Markets with Informational Frictions and Heterogeneous Production Technologies	6	06/03
807	P. Frijters M. A. Shields N. Theodoropoulos S. Wheatley Price	Testing for Employee Discrimination Using Matched Employer-Employee Data: Theory and Evidence	5	06/03
808	F. Docquier H. Rapoport	Remittances and Inequality: A Dynamic Migration Model	1	06/03
809	S. Commander M. Kangasniemi L. A. Winters	The Brain Drain: Curse or Boon?	1	06/03
810	J. H. Abbring G. J. van den Berg	A Simple Procedure for the Evaluation of Treatment Effects on Duration Variables	6	06/03

An updated list of IZA Discussion Papers is available on the center's homepage www.iza.org.