On Nonparametric Identification of Treatment Effects in Duration Models

Per Johansson
Myoung-jae Lee

July 2016

DISCUSSION PAPER SERIES

On Nonparametric Identification of Treatment Effects in Duration Models

Per Johansson
*IFAU, Uppsala University
and IZA*

Myoung-jae Lee
Korea University

ABSTRACT

On Nonparametric Identification of Treatment Effects in Duration Models

We show that the main nonparametric identification finding of Abbring and Van den Berg (2003b, Econometrica) for the effect of a timing-chosen treatment on an event duration of interest does not hold. The main problem is that the identification is based on the competing-risks identification result of Abbring and Van den Berg (2003a, Journal of the Royal Statistical Society, Series B) that requires independence between the waiting duration until treatment and the event duration, but the independence assumption does not hold unless there is no treatment effect. We illustrate the problem using constant hazards (i.e., exponential distribution), and as it turns out, there is no constant-hazard data generating process satisfying the assumptions in Abbring and Van den Berg (2003b, Econometrica) so long as the effect is not zero. We also suggest an alternative causal model.

Corresponding author:

Myoung-jae Lee
Department of Economics
Korea University
Seoul 136-701
South Korea
E-mail: myoungjae@korea.ac.kr

1   Introduction

In a typical static treatment effect framework, there appear a treatment D and a response variable Y, and the effect of D on Y is found by $E(Y|D=1) - E(Y|D=0)$ with some covariates controlled. In real life, however, things can happen over time such that individuals choose when to take the treatment so that there appears a waiting duration W to treatment in addition to D and Y. In such a case, finding the effect of D on Y is not so straightforward because the treatment group changes over time with more individuals getting treated as time passes. For instance, an unemployed person chooses W to enrol in a job training D after the waiting time W, and the response variable Y is the unemployed duration from the baseline of losing the previous job.

Another example is that a diseased person chooses W to take a medicine D, and the response Y is the survival duration from the baseline of the disease onset. In cases like this, the essential problem is that the untreated tend to have shorter Y than the treated, because the untreated die before getting treated. Even if there is no genuine treatment effect, the simple fact that one should survive long enough to get treated makes the treatment look like having a Y-lengthening effect. One way to avoid this illusionary effect would be "to treat the dead".

Abbring and Van den Berg (2003b) have been influential for finding effects of a timing-chosen treatment on a response duration. As of this writing, Abbring and Van den Berg (2003b) have been cited 508 times in *Google Scholar*, and there are many studies using their results: Abbring and Van den Berg (2004), Abbring et al. (2005), Bergemann et al. (2011), De Graaf-Zijl et al. (2011), Gaure et al. (2012), Hogelund et al. (2010), Jahn and Rosholm (2013), Kyyrä (2010), Kyyrä et al. (2013), Lalive et al. (2005,2008), Melberg et al. (2010), Richardson and Van den Berg (2001,2013), Roed et al. (2006), Svarer (2011), Svarer and Verner (2008), Tatsiramos (2010), Van den Berg et al. (2004), Van den Berg and Vikström (2013) and Van Ours and Williams (2009,2012). In addition to these published studies, although not listed here, there are tens of papers unpublished yet also invoking the results in Abbring and Van den Berg (2003b). For brevity, we will abbreviate Abbring and Van den Berg (2003a,b) as *AV1*

and *AV2*, respectively.

With t indexing time, AV2 envisioned three hazards: a waiting time hazard $h_w(t)$ for W, an untreated hazard $h_0(t)$ for Y before treatment, and a treated hazard $h_1(t)$ for Y after treatment as the treatment can alter the hazard for Y. Their identification proof was done in two stages. The first stage nonparametrically identifies $h_w(t)$ and $h_0(t)$ using the competing risks framework in AV1 based on observing only $\min(W, Y)$, and the second stage nonparametrically identifies $h_1(t)$. A critical assumption for the first stage is the conditional independence between W and Y given time-constant covariates and unobserved heterogeneities. We will show in this paper that this independence assumption cannot hold unless the treatment has no effect, which invalidates the main nonparametric identification finding of AV2, unfortunately. We then suggest another causal model for nonparametric identification.

The rest of this paper is organized as follows. Section 2 presents our main point: the failure of the conditional independence between W and Y so that the results in AV1 cannot be invoked. Section 3 provides a specific example using the simplest possible setup (i.e., constant hazards) with no covariate nor unobserved heterogeneity: there is no data generating process (DGP) that satisfies the assumptions in AV2 and give the desired identification. Section 4 presents another causal model where nonparametric identification holds. Finally, Section 5 concludes.

2 Failure of Independence and False Untreated

2.1 Dependence between W and Y

Let $h_w(t) > 0 \; \forall t$ be the hazard for W at time t that may depend on covariates X and an unobserved heterogeneity ("frailty"). Then W can be generated by solving

$$\exp\{-\int_0^W h_w(\tau)d\tau\} = U_w \tag{2.1}$$

where $U_w \sim U[0,1]$ with $U[0,1]$ denoting the uniform distribution over $[0,1]$. Since $\int_0^W h_w(\tau)d\tau$ is strictly increasing in W, the unique solution exists in (2.1), where the left-hand side is the survival function of W with W as its argument.

Let $h_0(t) > 0 \; \forall t$ be the hazard before treatment, and $h_1(t) > 0$ after treatment; $h_0(t)$ and $h_1(t)$ may also depend on X and unobserved heterogeneities, although the presence of these terms is suppressed for notational simplicity as they do not matter for our point. The AV2 causal model generates (W, Y) as follows. First, generate W using (2.1). Second, given W, generate Y by solving for Y

$$\exp\{-\int_0^W h_0(\tau)d\tau - \int_W^Y h_1(\tau)d\tau\} = U_y \tag{2.2}$$

where $U_y \sim U[0,1]$ and $U_y \amalg U_w$. AV2 (2003b, p. 1496) explained this Y-generating process using the fact $\Theta(M) \sim Expo(1)$ where Θ is the cumulative hazard function of a random variable M and $Expo(1)$ stands for exponential distribution with parameter 1, which is equivalent to using the survival function and $U[0,1]$ in (2.2).

The main identification finding of AV2 (pp. 1505-1506) is proved in two stages. The first stage invokes the identification result of AV1 for competing risks where only $\min(W, Y)$ is observed and two hazards for W and Y are identified nonparametrically, and the second stage refers to the case of both (W, Y) observed due to $W < Y$. *The first-stage identification result of AV1 (p. 702) assumes the conditional independence of W and Y given X and the unobserved heterogeneities. But this is violated in (2.2) because Y depends on W.*

It is curious why the above failure in the first-stage proof has not been noticed in the literature. The most likely reason is that AV2 is a difficult paper to read and the applied studies in the literature simply took the finding at the face value. It is curious that Gaure et al. (2007) used only zero-effect models for their simulation study to show that their estimator works; why not show this under a non-zero effect? Gaure et al. (2007) then used a model with non-zero heterogeneous effects only to demonstrate biases in their Table 12.

When does the problem of Y depending on W disappear? This happens if $h_0 = h_1$, under which (2.2) becomes

$$\exp\{-\int_0^Y h_0(\tau)d\tau\} = U_y$$

regardless of W. The two hazards h_w and h_0 are nonparametrically identified as AV1

showed because the conditional independence holds between W and Y generated by

$$\exp\{-\int_0^W h_w(\tau)d\tau\} = U_w \quad \text{and} \quad \exp\{-\int_0^Y h_0(\tau)d\tau\} = U_y \quad \text{with } U_w \amalg U_y. \quad (2.3)$$

2.2 False Untreated: Treating the Dead

It is important to be aware that *the post-treatment hazard h_1 and W are involved in generating Y in (2.2)* so long as $h_0 \neq h_1$. Hence so long as $h_0 \neq h_1$, any Y is a treated duration. Still yet, $Y < W$ (untreated) can happen if U_y in (2.2) is relatively large to make $\int_0^W h_0(\tau)d\tau + \int_W^Y h_1(\tau)d\tau$ small, which in turn forces Y small. Since W was generated using $\int_0^W h_w(\tau)d\tau$ in (2.1) whereas $\int_0^W h_0(\tau)d\tau$ appears for Y in (2.2), $Y < W$ can happen despite that h_1 and W are involved in generating Y. This is treating the dead: a person gets treated at W, but the person was in fact dead before W. Other than the false control group, AV2 presumed the "genuine" control group: those with $W = \infty$ (p. 1495).

Abbring and Van den Berg (2014, personal communication; *AV3* hereafter) acknowledged that Y and W are not independent. They, however, maintained that their finding still holds, because the subdensity functions of 'Y with $Y < W$' and 'W with $W < Y$' are, omitting the covariates and unobserved heterogeneities,

$$
\begin{aligned}
(i) \quad &: \quad h_0(y)\exp\{-\int_0^y h_0(\tau)d\tau\} \cdot \exp\{-\int_0^y h_w(\tau)d\tau\}, \quad &(2.4) \\
(ii) \quad &: \quad h_w(w)\exp\{-\int_0^w h_w(\tau)d\tau\} \cdot \exp\{-\int_0^w h_0(\tau)d\tau\}.
\end{aligned}
$$

AV3 stated "these are the subdensities of an independent competing risks model in which the risks have hazard rates h_w and h_0. This model does not involve the treatment effect, so its analysis can proceed without knowledge of h_1". The next section will use constant hazards

$$h_w(t) = \alpha > 0, \quad h_0(t) = \beta_0 > 0 \quad \text{and} \quad h_1(t) = \beta_1 > 0 \quad \forall t \quad (2.5)$$

to show that there is no constant-hazard-based DGP satisfying the assumptions in AV2, and that (2.4)(i) is false for feasible constant-hazard DGP's that would arise in real life.

A reviewer suggested another DGP that is free of the treating-the-dead problem:

$$(i) \quad : \quad \exp\{-\int_0^W h_0(\tau)d\tau - \int_W^Y h_1(\tau)d\tau\} = U_y \quad \text{if} \quad U_y \le \exp\{-\int_0^W h_0(\tau)d\tau\}$$

$$(ii) \quad : \quad \exp\{-\int_0^Y h_0(\tau)d\tau\} = U_y \quad \text{if} \quad U_y > \exp\{-\int_0^W h_0(\tau)d\tau\}. \quad (2.6)$$

To ease referring, call this 'DGP$_2$' as there are two equations, relative to 'DGP$_1$' in (2.2) with a single equation. For the constant hazards in (2.5), this becomes, with $\beta_d \equiv \beta_1 - \beta_0$,

$$(i) \quad : \quad \exp\{-\beta_0 W - \beta_1(Y-W)\} = U_y \iff Y = \frac{\beta_d}{\beta_1}W - \frac{1}{\beta_1}\ln U_y \quad \text{if} \quad U_y \le \exp(-\beta_0 W)$$

$$(ii) \quad : \quad \exp(-\beta_0 Y) = U_y \iff Y = -\frac{1}{\beta_0}\ln U_y \quad \text{if} \quad U_y > \exp(-\beta_0 W). \quad (2.7)$$

In DGP$_2$, there is no treating-the dead problem because $W \le Y$ in (i) and $Y < W$ in (ii); h_1 and W are involved in generating Y only in the former.

We think DGP$_2$ will give (2.4) even when $h_0 \ne h_1$; i.e., DGP$_2$ will pass the first stage to identify h_w and h_0. This would be a good news, but as we showed in the appendix, AV2 adopted DGP$_1$ instead of DGP$_2$. More importantly, even if DGP$_2$ is used for AV2, the second stage identification of h_1 fails due to the sample selection problem that U_y in (2.6)(i) is subject to $U_y \le \exp(-\beta_0 W)$ and thus related to W. The easiest way to see this selection problem is considering the least squares estimator (LSE) of Y on $(1, W)$ in (2.7)(i) to estimate β_d/β_1 where the error term is $-\beta_1^{-1}\ln U_y$: since U_y is related to W, the LSE is inconsistent. No instrument for W will help, because any variable affecting W would be related to $U_y|\{U_y \le \exp(-\beta_0 W)\}$.

There is a trade-off between the treating-the-dead and sample selection problems, which can be seen in comparing DGP$_1$ and DGP$_2$. DGP$_1$ avoids the selection problem because U_y is not related to W in DGP$_1$, but it suffers from the treating-the-dead problem. Viewed differently, by treated the dead, we can avoid the selection problem of the untreated having shorter durations than the treated simply because the untreated did not make it to the treatment even if the effect is zero. DGP$_2$ overcomes the treating-the-dead problem at the expense of the selection problem. DGP$_1$ fails in the first stage of the AV2 identification steps, whereas DGP$_2$ fails in the second stage. Notice that

DGP_2 includes DGP_1 as a limiting case with $W \to 0$: everybody treated except those with $W = \infty$.

3   DGP Problem under Constant Hazards

Suppose that constant hazards hold as in (2.5) to get

$$H_W(t) = \alpha t, \qquad f_W(w) = \alpha \exp(-\alpha w), \qquad F_W(w) = 1 - \exp(-\alpha w) \qquad (3.1)$$

where H_W denotes the cumulative hazard, f_W the density and F_W the distribution function. Then, recalling the treatment effect parameter $\beta_d \equiv \beta_1 - \beta_0$,

$$
\begin{aligned}
\exp(-\alpha W) &= U_w \implies W = -\frac{1}{\alpha} \ln U_w, \\
\exp\{-\beta_0 W - \beta_1 (Y - W)\} &= U_y \implies Y = \frac{\beta_d}{\beta_1} W - \frac{1}{\beta_1} \ln U_y.
\end{aligned}
\qquad (3.2)
$$

In (3.2), there is no guarantee for $Y > 0$ unless $\beta_d \geq 0$. Hence we will *assume* $\beta_d \geq 0$. We will *also assume* $\beta_d \neq \alpha$ as $(\beta_d - \alpha)^{-1}$ will appear below; this is only for brevity, as we can easily allow for the case $\beta_d = \alpha$.

For the constant hazards, the claimed subdensity function of Y with $Y < W$ by AV3 in (2.4)(i) equals

$$\beta_0 \exp\{-(\alpha + \beta_0)y\}. \qquad (3.3)$$

We will show in the next section that the true subdensity of Y with $Y < W$ equals the claimed form (3.3) under no effect $\beta_d = 0$, but otherwise not when $\beta_d > 0$.

3.1   Subdensity under Zero Effect

Suppose $\beta_d = 0$. Using (3.2) with $\beta_d = 0 \iff \beta_0 = \beta_1$, it holds that

$$P(Y > y | W = w) = P(-\frac{1}{\beta_0} \ln U_y > y | W = w).$$

'$|W = w$' drops out to result in $Y|(W = w) \sim Expo(\beta_0)$:

$$P(Y > y | W = w) = P(Y > y) = \exp(-\beta_0 y).$$

Consider

$$P(Y > y | Y < W) = \frac{P(Y > y, Y < W)}{P(Y < W)} = \frac{\text{'cause-}Y \text{ subsurvival function'}}{P(Y \text{ observed, not } W)}.$$

The appendix shows that

$$(i) \quad : \quad P(Y > y, Y < W) = \frac{\beta_0}{\alpha + \beta_0} \exp\{-(\alpha + \beta_0)y\} \tag{3.4}$$

$$(ii) \quad : \quad P(Y < W) = \frac{\beta_0}{\alpha + \beta_0}.$$

Hence, $P(Y > y | Y < W)$ follows $Expo(\alpha + \beta_0)$:

$$P(Y > y | Y < W) = \frac{\alpha + \beta_0}{\beta_0} \frac{\beta_0}{\alpha + \beta_0} \exp\{-(\alpha + \beta_0)y\} = \exp\{-(\alpha + \beta_0)y\}. \tag{3.5}$$

By symmetry, we also have $W | W < Y$ following $Expo(\alpha + \beta_0)$. That is, $\alpha + \beta_0$ is identified by both $Y | Y < W$ and $W | W < Y$. Combining this with $P(Y < W) = \beta_0/(\alpha + \beta_0)$, we can identify both α and β_0. Differentiating the subsurvival function (3.4)(i) with respect to y and then multiplying by -1, we obtain the true subdensity that equals the claimed form in (3.3) under no effect.

3.2   Subdensity under Positive Effect

Now suppose $\beta_d > 0$ and $\beta_d \neq \alpha$. The appendix proves that

$$(i) \quad : \quad P(Y > y, Y < W)$$

$$= \frac{\beta_d}{(\beta_d - \alpha)} \exp(-\frac{\alpha}{\beta_d}\beta_1 y) - \frac{\alpha\beta_1}{(\beta_d - \alpha)(\alpha + \beta_0)} \exp\{-(\alpha + \beta_0)y\} \tag{3.6}$$

$$(ii) \quad : \quad P(Y < W) = \frac{\beta_0}{\alpha + \beta_0}.$$

The ratio of (i) to (ii) is

$$P(Y > y | Y < W) = \frac{(\alpha + \beta_0)\beta_d}{\beta_0(\beta_d - \alpha)} \exp(-\frac{\alpha}{\beta_d}\beta_1 y) - \frac{\alpha\beta_1}{\beta_0(\beta_d - \alpha)} \exp\{-(\alpha + \beta_0)y\}; \tag{3.7}$$

if $\beta_d = 0 \iff \beta_0 = \beta_1$, (3.7) becomes $\exp\{-(\alpha + \beta_0)y\}$ in (3.5). Note that (3.7) converges to 0 as $y \to \infty$, and converges to 1 as $y \to 0$ because

$$\frac{(\alpha + \beta_0)\beta_d}{\beta_0(\beta_d - \alpha)} - \frac{\alpha\beta_1}{\beta_0(\beta_d - \alpha)} = \frac{\alpha\beta_d + \beta_0\beta_d - \alpha(\beta_0 + \beta_d)}{\beta_0(\beta_d - \alpha)} = \frac{\beta_0(\beta_d - \alpha)}{\beta_0(\beta_d - \alpha)} = 1.$$

Differentiating the subsurvival function in (3.6)(i) with respect to y and then multiplying by -1, we obtain the subdensity for Y with $Y < W$:

$$\frac{\alpha\beta_1}{(\beta_d - \alpha)}\{\exp(-\frac{\alpha}{\beta_d}\beta_1 y) - \exp\{-(\alpha + \beta_0)y\}. \tag{3.8}$$

Note that the difference between the two exponents is

$$-\frac{\alpha}{\beta_d}\beta_1 + (\alpha + \beta_0) = \frac{1}{\beta_d}(-\alpha\beta_1 + \alpha\beta_d + \beta_0\beta_d) = \frac{1}{\beta_d}(-\alpha\beta_0 + \beta_0\beta_d) = \frac{\beta_0}{\beta_d}(\beta_d - \alpha) :$$

this is positive/negative according to $\beta_d - \alpha$ positive/negative, which ensures that the subdensity is positive always. *The subdensity function (3.8) for Y with $Y < W$ differs from the claimed form (3.3), and (3.8) does involve β_d.*
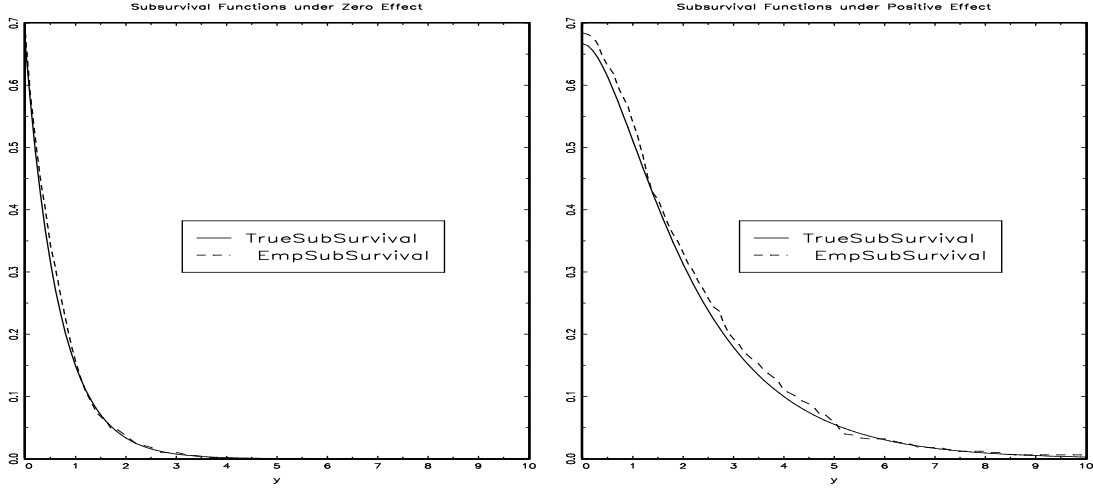

Figure 1: True & Empirical SubSurvivial Functions under Zero & Positive Effects

To dissipate any doubt the reader might have about (3.6), we generated data using

$$N = 500, \qquad \alpha = 0.5, \qquad \beta_0 = 1, \qquad \beta_d = 0 \text{ or } 5$$

and plotted the true subsurvival functions $y \longmapsto P(Y > y, Y < W)$ in solid lines and its estimated versions in dashed lines. The left panel of Figure 1 is for no effect ($\beta_d = 0$), whereas the right panel is for a positive effect ($\beta_d = 5$). Figure 1 demonstrates that the empirical subsurvival functions closely match the true versions in (3.4)(i) and (3.6)(i). Also $P(Y < W) = 0.67$ is closely matched by two sample means 0.65 (under $\beta_d = 0$) and 0.68 (under $\beta_d = 5$).

9

Equation (A.2) in the appendix shows

$$P(Y > y|W = w) = 1[y \le \frac{\beta_d}{\beta_1}w] \; + \; 1[\frac{\beta_d}{\beta_1}w < y]\exp\{-\beta_0 y - \beta_d(y - w)\}.$$

Taking ln on this survival function and multiplying by -1 gives the cumulative hazard:

$$0 \text{ when } y \le \frac{\beta_d}{\beta_1}w \quad \text{and} \quad \beta_0 y + \beta_d(y - w) \text{ when } \frac{\beta_d}{\beta_1}w < y.$$

Consider now $P(Y > y|W = w)$ and $P(Y > y|W = w')$ with $w \ne w'$. Their cumulative hazards do not fully agree on $y \in \{0, \min(w, w')\}$ due to $\beta_d(y - w)$ versus $\beta_d(y - w')$. This violates the no-anticipation assumption of AV2 (p. 1497), as the anticipation behavior is evident in $\beta_d(y - w)$ and $\beta_d(y - w')$ for the intervention timing. Hence, the no-anticipation assumption rules out $\beta_d > 0$.

With $\beta_d < 0$ already ruled out as a DGP, there is no DGP for AV2 under constant hazards and $\beta_d \ne 0$. This casts doubt on the role of the no-anticipation assumption, because the constant hazard model with $\beta_d > 0$ is the simplest DGP one can think of that can easily arise in reality. In short, there is nothing left, as far as DGP under constant hazards and non-zero effect goes.

4 Causal Model with Three Durations

Having presented a negative result, one may wonder if it is possible to have a nonparametric identification for the effect of a timing-chosen treatment on a response duration. The answer is positive when the 'three-duration' framework in Lee and Johansson (2013) is adopted.

Lee and Johansson (2013) considered three durations: W, an untreated duration Y^0 and a treated duration Y^1, instead of the two durations (W, Y). In the three-duration framework, W and Y^0 are generated with

$$\exp\{-\int_0^W h_w(\tau)d\tau\} = U_w \quad \text{and} \quad \exp\{-\int_0^{Y^0} h_0(\tau)d\tau\} = U_0 \quad (4.1)$$
$$\text{where } U_w \text{ and } U_0 \text{ are iid } U[0, 1].$$

Then the 'net treated duration' Y^1 from the time W and onward is generated by

$$\exp\{-\int_0^{Y^1} h_1(\tau)d\tau\} = U_1 \quad \text{where } U_1 \sim U[0, 1]; \quad (4.2)$$

$h_1(t)$ may depend on W to make Y^1 depend on W, in which case 'Y^w' would be more appropriate as the potential treated duration when treated at w. The 'gross treated duration' from the baseline is

$$W + Y^1 \quad (\text{or } W + Y^W). \tag{4.3}$$

Nonparametric identification of the three hazards can be done in two stages laid out in the next paragraph, under the assumption of mixed proportional hazards (MPH) and the assumption that (Y^1, Y^0, W) are conditionally independent given X and the unobserved heterogeneity factors. The MPH assumption is that each hazard is a product of a function of t (the baseline hazard), a function of X, and an unobserved factor; the three unobserved factors are assumed to be independent of X, although they may be related to one another. AV1 and AV2 required the MPH assumption. All functional forms including the distribution of the unobserved factors are allowed to be unknown.

In the first stage, both h_w and h_0 are nonparametrically identified due to AV1's competing-risks identification finding, because W and Y^0 are conditionally independent given X and the unobserved factors. In the second stage, h_1 is nonparametrically identified using only the treatment group (i.e., the group with $W < Y^0$ so that W is observed) due to the assumption that (Y^1, Y^0, W) are conditionally independent given X and the unobserved factors. This is so because the selection problem of conditioning on $W < Y^0$ does not occur under the conditional independence assumption, and because identifying h_1 using only Y^1 is strictly easier than the first stage identification.

5 Conclusions

We showed that the critical independence assumption between treatment-waiting and response durations invoked in Abbring and Van den Berg (2003b) fails, and thus their main nonparametric identification finding does not hold unless there is no treatment effect. We also showed that, under constant hazards, there is no DGP satisfying the assumptions of Abbring and Van den Berg (2003b) due to the critical no anticipation assumption. A causal model based on three durations was then presented as an alternative where nonparametric identification holds.

Will there be any valid DGP for non-constant hazards? Perhaps, but unlikely, as constant hazards are the simplest and they are the building blocks for more sophisticated hazards. Certainly, we cannot prove that no valid DGP exists for any hazards; rather, it will be up to Abbring and Van den Berg to demonstrate that a practically reasonable and useful DGP holds within their framework, e.g., under the popular Weibull hazards. Abbring and Van den Berg (2003b) did not provide any specific DGP that satisfies all their assumptions, although they provided a specific example where the identification fails. Our conjecture is that there exists no such DGP. This may be the reason why there has been no simulation evidence in the literature that their identification and ensuing estimation work.

Despite the negative points just mentioned, this paper also provided a positive point: if a different DGP without the "treating-the-dead" problem is adopted that a reviewer suggested and if the selection problem associated with the DGP can be dealt with somehow, then the nonparametric identification of Abbring and Van den Berg (2003b) may hold. Of course, it would not be easy to overcome the selection issue, but at least there is a direction to pursue.

APPENDIX

DGP_1 Adopted by AV2

With our W being their S in AV2, AV2 (p. 1495) stated their causal model as

> "our dynamically assigned binary treatment can be reinterpreted as a set of mutually exclusive treatment indexed by $\bar{R}_+ \equiv R_+ \cup \{\infty\}$. Here, the point ∞ represents the no-treatment case. To each treatment $s \in \bar{R}_+$ corresponds a random variable $Y^*(s) \geq 0$, the potential outcome in the case that we would intervene and assign treatment s.",......,"The actual outcome is $Y := Y^*(S)$;"

Hence, in AV2, W is a mixture of a R_+-valued random variable and ∞. That is, generate a binary random variable Z, and if $Z = 0$, set $W = \infty$; otherwise, generate

W from, e.g., $Expo(\alpha)$. The genuine control group $(Z = 0)$ occurs only because they get never treated with $W = \infty$. For those with $Z = 1$, the intervention at w alters the hazard from h_0 to h_1 to give

$$h_{Y^*(w)}(y) = h_0(y)1[y \leq w] + h_1(y)1[w < y];$$

Model 1A of AV2 (p.1503) is a sophisticated version of this. This leads to defining $Y^*(w)$ by inverting the cumulative hazard function in

$$\exp\{-\int_0^w h_0(\tau)d\tau - \int_w^{Y^*(w)} h_1(\tau)d\tau\} = U_y;$$

see the equation in the middle of AV2 (p. 1496). Replacing w with W in this display gives the observed $Y \equiv Y^*(W)$.

In discussing their Model 1A, AV2 (p. 1504, lines 4-5) stated "Conditional on (X, V), the variables Y and S are only dependent through $\delta(t|S, X)$", where $\delta(t|S, X)$ is the hazard shifter. DGP_2 goes against this, because the hazard shifter is a constant in the constant hazard model, and yet Y and W are dependent through the selection condition. Also, it should be noted that AV3 did not object to DGP_1, nor did they suggest DGP_2.

Proof for (3.4)

Observe

$$P(Y > y, Y < W) = \int P(Y > y, Y < w|W = w)f_W(w)dw \qquad (A.1)$$
$$= \int P(y < Y < w)1[y < w]\alpha \exp(-\alpha w)dw$$

$$= \alpha \int_y^\infty \{\exp(-\beta_0 y) - \exp(-\beta_0 w)\} \exp(-\alpha w)dw$$
$$= \exp(-\beta_0 y)\alpha \int_y^\infty \exp(-\alpha w)dw - \alpha \int_y^\infty \exp\{-(\alpha + \beta_0)w\}dw]$$
$$= \exp(-\beta_0 y)\exp(-\alpha y) - \frac{\alpha}{\alpha + \beta_0} \exp\{-(\alpha + \beta_0)y\}$$
$$= \exp\{-(\alpha + \beta_0)y\} - \frac{\alpha}{\alpha + \beta_0} \exp\{-(\alpha + \beta_0)y\} = \frac{\beta_0}{\alpha + \beta_0} \exp\{-(\alpha + \beta_0)y\}.$$

Also observe

$$P(Y < W) = \int P(Y < w|W = w)f_W(w)dw = \alpha \int \{1 - \exp(-\beta_0 w)\}\exp(-\alpha w)dw$$

$$= \alpha \int_0^\infty \exp(-\alpha w)dw - \alpha \int_0^\infty \exp\{-(\alpha + \beta_0)w\}dw = 1 - \frac{\alpha}{\alpha + \beta_0} = \frac{\beta_0}{\alpha + \beta_0}.$$

Proof for (3.6)

As a preliminary, we have

$$P(Y > y|W = w) = P(\frac{\beta_d}{\beta_1}W - \frac{1}{\beta_1}\ln U_y > y|W = w) = P\{U_y < \exp(-\beta_1 y + \beta_d w)\}$$

$$= \exp(-\beta_1 y + \beta_d w) \quad \text{if} \ \exp(-\beta_1 y + \beta_d w) < 1 \quad \text{and} \ 1 \ \text{otherwise}$$

$$= \exp(-\beta_1 y + \beta_d w) \cdot 1[w < \frac{\beta_1}{\beta_d}y] \ + \ 1[\frac{\beta_1}{\beta_d}y \le w]. \tag{A.2}$$

It is interesting that despite $\beta_d \ne 0$, $y = w$ gives, since $\beta_1/\beta_d > 1 \implies w < \beta_1 w/\beta_d$,

$$P(Y > w|W = w) = \exp(-\beta_1 w + \beta_d w) = \exp(-\beta_0 w) \tag{A.3}$$

as if Y were independent of W and $Y \sim Expo(\beta_0)$ due to $\beta_d = 0$.

Recalling (A.1) and (A.2), it holds that

$$P(Y > y, Y < W) = \int P(y < Y < w|W = w)1[y < w]\alpha \exp(-\alpha w)dw$$

$$= \alpha \int_y^\infty \{P(y < Y|W = w) - P(w < Y|W = w)\}\exp(-\alpha w)dw$$

$$= \alpha \int_y^\infty \{\exp(-\beta_1 y + \beta_d w)1[w < \beta_d^{-1}\beta_1 y]$$

$$+ 1[\beta_d^{-1}\beta_1 y \le w] - \exp(-\beta_0 w)\}\exp(-\alpha w)dw$$

$$= \alpha \exp(-\beta_1 y)\int_y^{\beta_d^{-1}\beta_1 y}\exp\{(\beta_d - \alpha)w\}dw$$

$$+ \alpha \int_{\beta_d^{-1}\beta_1 y}^\infty \exp(-\alpha w)dw - \alpha \int_y^\infty \exp\{-(\alpha + \beta_0)w\}dw$$

$$= \frac{\alpha}{\beta_d - \alpha}\exp(-\beta_1 y)\exp\{(\beta_d - \alpha)w\}|_y^{\beta_d^{-1}\beta_1 y} \tag{A.4}$$

$$+ \exp(-\alpha \beta_d^{-1}\beta_1 y) - \frac{\alpha}{\beta_0 + \alpha}\exp\{-(\alpha + \beta_0)y\}.$$

The first term in (A.4) equals

$$\frac{\alpha}{\beta_d - \alpha}\exp(-\beta_1 y)[\exp\{(\beta_d - \alpha)\beta_d^{-1}\beta_1 y\} - \exp\{(\beta_d - \alpha)y\}]$$

$$= \frac{\alpha}{\beta_d - \alpha}[\exp(-\alpha \beta_d^{-1}\beta_1 y) - \exp\{-(\alpha + \beta_0)y\}].$$

14

Substituting this into (A.4), we obtain (3.6)(i):

$$\frac{\alpha}{\beta_d - \alpha}[\exp(-\alpha\beta_d^{-1}\beta_1 y) - \exp\{-(\alpha + \beta_0)y\}]$$
$$+ \exp(-\alpha\beta_d^{-1}\beta_1 y) - \frac{\alpha}{\beta_0 + \alpha}\exp\{-(\alpha + \beta_0)y\}$$
$$= \frac{\beta_d}{(\beta_d - \alpha)}\exp(-\frac{\alpha}{\beta_d}\beta_1 y) - \frac{\alpha\beta_1}{(\beta_d - \alpha)(\alpha + \beta_0)}\exp\{-(\alpha + \beta_0)y\}.$$

As for (3.6)(ii), it holds due to (A.3):

$$P(Y < W) = \int P(Y < w|W = w)f_W(w)dw = \alpha \int \{1 - \exp(-\beta_0 w)\}\exp(-\alpha w)dw.$$

REFERENCES

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

Abbring, J.H. and G.J. Van den Berg, 2003b, The nonparametric identification of treatment effects in duration models, Econometrica 71, 1491-1518.

Abbring, J.H. and G.J. Van den Berg, 2004, Analyzing the effect of dynamically assigned treatments using duration models, binary treatment models, and panel data models, Empirical Economics 29, 5-20.

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

Bergemann, A.H., M. Caliendo, G.J. van den Berg and K.F. Zimmermann, 2011, The threat effect of participation in active labor market programs on job search behavior of migrants in Germany, International Journal of Manpower 32, 777-795.

De Graaf-Zijl, M., G.J. van den Berg and A. Heyma, 2011, Stepping stones for the unemployed: the effect of temporary jobs on the duration until (regular) work, Journal of Population Economics 24, 107-139.

Gaure, S., K. Roed and T. Zhang, 2007, Time and causality: a Monte Carlo assessment of the timing-of-events approach, Journal of Econometrics 141, 1159-1195.

Gaure, S., K. Roed and L. Westlie, 2012, Job search incentives and job match quality, Labour Economics 19, 438-450.

Hogelund, J., A. Holm and J. McIntosh, 2010, Does graded return-to-work improve sicklisted workers' chance of returning to regular working hours?, Journal of Health Economics 29, 158-169.

Jahn, E. and M. Rosholm, 2013, Is temporary agency employment a stepping stone for immigrants?, Economics Letters 118, 225-228.

Kyyrä, T., 2010, Partial unemployment insurance benefits and the transition rate to regular work, European Economic Review 54, 911-930.

Kyyrä, T., P. Parrotta and M. Rosholm, 2013, The effect of receiving supplementary UI benefits on unemployment duration, Labour Economics 21, 122-133.

Lalive, R., J.C. van Ours and J. Zweimüler, 2005, The effect of benefit sanctions on the duration of unemployment Journal of the European Economic Association 3, 1386-1417.

Lalive, R., J.C. van Ours and J. Zweimüler, 2008, The impact of active labor market programs and benefit entitlement rules on the duration of unemployment, Economic Journal 118, 235-257.

Lee, M.J. and P. Johansson, 2013, Three-duration approach for treatment effects on durations when the treatment timing is chosen, presented at the 67th European Meeting of the Econometric Society, Gothenburg, Sweden (an earlier version in IZA Discussion Paper 7249, 2013 February).

Melberg, H.O., A. Jones and A. Bretteville-Jensen, 2010, Is cannabis a gateway to hard drugs?, Empirical Economics 38, 583-603.

Richardson, K. and G.J. van den Berg, 2001, The effect of vocational employment training on the individual transition rate from unemployment to work, Swedish Economic Policy Review 8, 175-213.

Richardson, K. and G.J. van den Berg, 2013, Duration dependence versus unobserved heterogeneity in treatment effects: Swedish labor market training and the

transition rate to employment, Journal of Applied Econometrics 28, 325-351.

Roed, K. and O. Raaum, 2006, Do labour market programmes speed up the return to work?, Oxford Bulletin of Economics and Statistics 68, 541-568.

Svarer, M., 2011, Crime and partnerships, Review of Economics of the Household 9, 307-325.

Svarer M. and M. Verner, 2008, Do children stabilize relationships in Denmark?, Journal of Population Economics 21, 395-417l.

Tatsiramos, K., 2010, Job displacement and the transitions to re-employment and early retirement for non-employed older workers, European Economic Review 54, 517-535.

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, 211-241.

Van den Berg, G.J. and J. Vikström, 2013, Monitoring job offer decisions, punishments, exit to work, and job quality, Scandinavian Journal of Economics, forthcoming.

Van Ours, J.C. and J. Williams, 2009, Why parents worry: initiation into cannabis use by youth and their educational attainment, Journal of Health Economics 28, 132-142.

Van Ours, J.C. and J. Williams, 2012, The effects of cannabis use on physical and mental health, Journal of Health Economics 31, 564-577.