Rebuttal of „On Nonparametric Identification of Treatment Effects in Duration Models"

Jaap H. Abbring
Gerard J. van den Berg

Rebuttal of "On Nonparametric Identification of Treatment Effects in Duration Models"

Jaap H. Abbring
CentER, Tilburg University and CEPR

Gerard J. van den Berg
University of Bristol,
IFAU, IZA, ZEW, University of Mannheim and CEPR

ABSTRACT

Rebuttal of "On Nonparametric Identification of Treatment Effects in Duration Models"[*]

In their IZA Discussion Paper 10247, Johansson and Lee claim that the main result (Proposition 3) in Abbring and Van den Berg (2003b) does not hold. We show that their claim is incorrect. At a certain point within their line of reasoning, they make a rather basic error while transforming one random variable into another random variable, and this leads them to draw incorrect conclusions. As a result, their paper can be discarded.

JEL Classification: C14, C31, C41

Keywords: econometrics, survival analysis, hazard rates, scientific conduct

Corresponding author:

Gerard J. van den Berg
School of Economics, Finance and Management
University of Bristol
2C3, The Priory Road Complex
Priory Road, Clifton
Bristol, BS8 1TU
United Kingdom
E-mail: gerard.vandenberg@bristol.ac.uk

1 Introduction

In a 2003 article in Econometrica (Abbring and Van den Berg, 2003b) we analyzed the specification and identification of causal multivariate duration models. We focused on the case in which one duration captures the point in time at which a treatment is initiated and one is interested in the effect of this treatment on some outcome duration. We defined a "no anticipation of treatment" assumption and showed that this assumption in a model framework inspired by the mixed proportional hazards model delivers identification. The model framework allows for dependent unobserved heterogeneity and does not require exclusion restrictions on covariates. In a nutshell, the timing of events conveys useful information on the treatment effect.

In their IZA Discussion Paper 10247, Johansson and Lee (henceforth JL) claim that our main identification result (Proposition 3, for our Model 1A) does not hold.[1] JL examine two model versions, which they refer to as DGP1 and DGP2. It is good custom that comments on articles from other researchers phrase their statements in terms of the notation and model specification of the original article. However, JL do not make this effort. They claim that their DGP1 captures our model (in particular, our Model 1A) and that our proof of Proposition 3 does not apply to DGP1. From this, they conclude that our proof is incorrect and Proposition 3 is not true.

In this note we show that JL's claims are incorrect. They make a rather basic error when deriving their DGP1 from our Model 1A. Consequently, DGP1 and our Model 1A are not equivalent, and the fact that our proof of Proposition 3 does not apply to DGP1 has no bearing on its validity. We show that, in fact, their DGP2, for which they do not raise similar concerns, is consistent with our model, in contrast to JL's claims.

In Section 2 we develop the setup in terms of JL's notation. We discuss properties of our Model 1A and show that these immediately refute statements early on in JL about our identification analysis. In Section 3 we explain the key error in JL's derivations. Section 4 concludes and critically assesses awkward remarks in JL about the empirical researchers who cited our 2003 Econometrica article.

---

[1]JL do not specifically refer to our Proposition 3. However, halfway their page 4, they do refer to the "main identification finding of [Abbring and Van den Berg (2003b)] (pp. 1505–1506)", and our Proposition 3 is the only relevant formal identification result on the pages 1505–1506 of the Econometrica article. Moreover, on page 13 of JL, they refer to our Model 1A, and Proposition 3 is the main identification result for Model 1A.

2   Johansson and Lee's setup and notation

The setup, in JL's notation, is a model with a treatment duration W and an outcome duration Y (throughout, following JL, we will keep observed covariates and unobserved heterogeneity implicit as these are not relevant to the argument). The treatment duration W has a hazard rate $h_W(w)$ and Lebesgue density

$$f_W(w) = h_W(w) \exp\left(-\int_0^w h_W(\tau)d\tau\right) \tag{1}$$

at duration w. The outcome duration Y given $W = w$ has a hazard rate $h_0(y)$ at times $y \leq w$ and a hazard rate $h_1(y)$ at times $y > w$, with corresponding Lebesgue density

$$f_{Y|W}(y|w) = \begin{cases} h_0(y) \exp\left(-\int_0^y h_0(\tau)d\tau\right) & \text{if } y \leq w \text{ and} \\ h_1(y) \exp\left(-\int_0^w h_0(\tau)d\tau - \int_w^y h_1(\tau)d\tau\right) & \text{if } y > w. \end{cases} \tag{2}$$

The joint density of (W, Y) at (w, y) is $f_{Y|W}(y|w)f_W(w)$.

In our 2003 Econometrica article, we were specifically interested in the differences between the hazard rates h_0 and h_1, which capture the "treatment effects" of the event at time W on the outcome duration Y. Therefore, we provided results on the identification of h_0 and h_1 under general conditions. To this end, we used that our Model 1A embeds an independent competing risks model with hazard rates h_W and h_0 for the so-called *"identified minimum"* of W and Y (i.e., their minimum *and* whether this minimum is equal to W or equal to Y). That is, we used that the subdensities of Y on $\{Y < W\}$ and W on $\{W < Y\}$ (which, together, fully characterize the distribution of this identified minimum) equal

$$-\frac{d}{dy}\Pr\left(Y > y, Y < W\right) = h_0(y)\exp\left(-\int_0^y h_0(\tau)d\tau - \int_0^y h_W(\tau)d\tau\right) \tag{3}$$

and

$$-\frac{d}{dw}\Pr\left(W > w, W < Y\right) = h_W(w)\exp\left(-\int_0^w h_W(\tau)d\tau - \int_0^w h_0(\tau)d\tau\right), \tag{4}$$

respectively (see also JL's (2.4)(i) and (2.4)(ii)). This intermediate result is key because it allowed us to apply Abbring and Van den Berg's (2003a) identification result for the competing risks model to establish the identification of h_W and h_0 (as well as, in the general setting with observed covariates and unobserved heterogeneity, the identification of the covariate effects on these hazard rates and the distribution of the unobserved heterogeneity

in them) from data on the identified minimum of W and Y. Abbring and Van den Berg (2003a) proved a one-to-one mapping between the competing risks model and the distribution of the identified minimum. As is intuitively clear, and apparent from Equations (3) and (4), the distribution of the identified minimum of W and Y only involves treatment and outcome hazards before either the treatment or the outcome event occurs and does not depend on the post-treatment outcome hazard (or, for that matter, the post-outcome treatment hazard). Indeed, the specification of the hazard rates after the minimum of W and Y is irrelevant for the competing risks identification result. The competing risks model in Abbring and Van den Berg (2003a) and the competing risks part of our Model 1A are equivalent.[2] In particular, in Model 1A (without covariates and unobserved heterogeneity), the implied submodel for the identified minimum of W and Y is an *independent* competing risks model, despite the fact that W and Y will generally be dependent if h_0 and h_1 differ.

JL state halfway their page 4 that $h_0 = h_1$ ("no treatment effects") is *necessary* for our key intermediate result that Model 1A embeds an independent competing risks model with hazard rates h_W and h_0. This is false. This should be clear from the previous paragraphs, but we can illuminate it further by examining the subsurvival functions $\Pr(Y > y, Y < W)$ of Y on $\{Y < W\}$ and $\Pr(W > w, W < Y)$ of W on $\{W < Y\}$, in order to verify that Equations (3) and (4) hold for general h_0 and h_1, including cases where $h_0 \neq h_1$. The subsurvival function of Y on $\{Y < W\}$ satisfies

$$
\begin{aligned}
&\Pr\left(Y > y, Y < W\right) \\
&= \int \Pr\left(Y > y, Y < W | W = w\right) f_W(w) dw \\
&= \int_y^\infty \Pr\left(y < Y < w | W = w\right) f_W(w) dw \\
&= \int_y^\infty \int_y^w f_{Y|W}(\tau|w) d\tau \, f_W(w) dw \\
&= \int_y^\infty \left[\exp\left(-\int_0^y h_0(\tau) d\tau\right) - \exp\left(-\int_0^w h_0(\tau) d\tau\right) \right] f_W(w) dw.
\end{aligned}
\tag{5}
$$

Differentiating the left-hand and right-hand sides of (5) with respect to y, substituting (1), integrating, and multiplying by -1 gives (3). An analogous derivation for the subsurvival function of W on $\{W < Y\}$ gives (4). Consequently, (3) and (4) hold generally, and this

---

[2]Underlying this equivalence is the no-anticipation assumption for potential outcomes made in Abbring and Van den Berg (2003b), which ensures that the potential outcome hazards before treatment do not depend on the eventual treatment time.

confirms that, in contrast to what JL claim halfway their page 4, Abbring and Van den Berg's (2003a) results for the competing risks model can be applied to the model for the identified minimum of W and Y embedded in Abbring and Van den Berg's (2003b) model of treatment effects.

3   Johansson and Lee's error

It is easy to see where JL go wrong in their own derivations. Halfway their page 6, they claim that we have adopted DGP1, which they associate with their Equation (2.2). For a proof, they refer to their Appendix. In this Appendix, at the top of page 13, they specify the hazard rate of the potential outcome $Y^*(w)$, evaluated at the elapsed duration y, as $h_0(y)$ if $y \leq w$ and as $h_1(y)$ if $y > w$. This hazard rate specification indeed corresponds to our Model 1A. The subsequent equation in their Appendix is supposed to capture $Y^*(w)$ as the inverse of the corresponding integrated hazard evaluated in a standard exponential random variable, as is clear from their reference to the equation halfway Abbring and Van den Berg (2003b, p. 1496) and the corresponding discussion below their Equation (2.2). However, the integrated hazard that is used here is not the correct integrated hazard, because it is not the integral of the hazard in the equation at the top of page 13. Specifically, if $w > Y^*(w)$ then the integrated hazard should equal $\int_0^{Y^*(w)} h_0(\tau)d\tau$. Instead, in this case, JL take it to equal $\int_0^w h_0(\tau)d\tau - \int_{Y^*(w)}^w h_1(\tau)d\tau$.[3] Of course, this leads to a very peculiar DGP1 with absurd implications. However, DGP1 is not consistent with our Model 1A and its absurd implications are solely due to the mistake by JL and do not carry over to our Model 1A.

If JL would have used the correct integrated hazard instead, they would have ended up with their DGP2 instead of DGP1. To see this, note that their Equation (2.6)(i) applies if

$$\exp\left[-\int_0^W h_0(\tau)d\tau - \int_W^Y h_1(\tau)d\tau\right] \leq \exp\left[-\int_0^W h_0(\tau)d\tau\right] \iff Y \geq W$$

and that their Equation (2.6)(ii) applies if

$$\exp\left[-\int_0^Y h_0(\tau)d\tau\right] > \exp\left[-\int_0^W h_0(\tau)d\tau\right] \iff Y < W,$$

---

[3]From JL's calculations for the special case with constant hazards, in particular their derivation of (3.6) in their Appendix, it is clear that that they indeed interpret $\int_w^{Y^*(w)} h_1(\tau)d\tau$ as $-\int_{Y^*(w)}^w h_1(\tau)d\tau$ if $w > Y^*(w)$.

where we have used JL's assumption that $h_0(\tau) > 0$ and $h_1(\tau) > 0$ for all τ (page 4). Moreover, the left-hand side of Equation (2.6)(i) involves the correct integrated hazard of Y evaluated at random W and Y such that $Y \geq W$, $\int_0^W h_0(\tau)d\tau + \int_W^Y h_1(\tau)d\tau$, and the left-hand side of Equation (2.6)(ii) involves the correct integrated hazard of Y evaluated at random W and Y such that $Y \leq W$, $\int_0^Y h_0(\tau)d\tau$ (note that both are identical and correct in the boundary case $Y = W$). Consequently, in contrast to what JL claim, their DGP2 specified by Equation (2.6) is consistent with our Model 1A. Moreover, as JL note halfway page 6, under DGP2, (3) and (4) hold for general h_0 and h_1. Taken together, this implies that JL's key claim that (3) and (4) cannot be used in the identification analysis of Model 1A is wrong.

Halfway page 6, JL raise a new concern about their DGP2, and thus our Model 1A: Information on the post-treatment outcome hazard h_1 can only be obtained from the selected subpopulation with $Y > W$. We share this concern; in fact, this is one aspect of the selection problem that is at the core of our paper and that we addressed successfully in it.[4] JL do not show that this concern invalidates our identification results, in particular our Proposition 3. Rather, they propose to infer some treatment effect parameters by regressing Y on W in a subsample of the population with $Y > W$ and claim that the resulting estimator is inconsistent. We never proposed this ad hoc procedure and we would certainly not recommend it. In any case, its failure to produce a consistent estimator of certain treatment effects does not prove our identification results wrong.

In sum, JL make a basic error in deriving their DGP1 from our Model 1A. This leads them to incorrectly claim that their DGP1 and our Model 1A are equivalent. Instead, our Model 1A corresponds their DGP2, to which their concerns about our identification analysis do not apply. After reading JL, one may wonder why these misunderstandings have not surfaced in an earlier stage of the publication process. Indeed, when presenting their claim that we adopted DGP1 in Abbring and Van den Berg (2003b), JL note halfway their page 14 that "[Abbring and Van den Berg] did not object to DGP1, nor did they suggest DGP2". Now, JL have not offered us the opportunity to respond to the most recent version of their note before they submitted it for publication as an IZA Discussion Paper. We did communicate with JL about previous drafts of their paper in 2014, both directly, after they sent us their paper's first draft in January 2014, and indirectly through the

---

[4]Other, more subtle aspects of this problem concern the selection on the unobservable heterogeneity factors that is kept implicit in JL and here.

editorial process at a journal. In these communications, we pointed out a logical flaw in an argument they used at the time, and that flaw has disappeared from the current draft (they claimed that $h_0 = h_1$ is necessary for (3) and (4) to hold but only showed it to be sufficient). We also directly demonstrated that their claim that $h_0 = h_1$ is necessary for (3) and (4) to hold is incorrect, by providing, as in this Rebuttal, elementary calculations that show that these equations hold generally. At the time, we did not provide an exhaustive list of all aspects of their paper that we disagreed with, because this was not necessary to make our point that they were wrong. In particular, we did not specifically refer to JL's DGP1 in our earlier private communication with them. The reader can rest assured, though, that both of us disagree with many things that we never explicitly mentioned or objected to.

4   Conclusion

The claims in JL's paper about the results in Abbring and Van den Berg (2003b) can be discarded.

In their paper, JL observe that our paper has been often cited by empirical studies, and they mention a number of authors who cited our work. In the light of what JL perceive as an incorrectness in our paper, they speculate about the reason for the high citation score. Specifically, they claim that "the most likely reason is that Abbring and Van den Berg (2003b) is a difficult paper to read and the applied studies in the literature simply took the finding at the face value". We view this as a preposterous statement. It offends the empirical researchers who cite our work, by depicting them as simple minds unable to understand methodological work. Our current note provides a better reason: our 2003 finding is correct.

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 non-parametric identification of treatment effects in duration models", *Econometrica*, 71, 1491-1517.