

IZA DP No. 7249

**A Simple Approach to Treatment Effects on Durations
When the Treatment Timing is Chosen**

Myoung-jae Lee
Per Johansson

February 2013

A Simple Approach to Treatment Effects on Durations When the Treatment Timing is Chosen

Myoung-jae Lee
Korea University

Per Johansson
*IFAU, Uppsala University
and IZA*

Discussion Paper No. 7249
February 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

A Simple Approach to Treatment Effects on Durations When the Treatment Timing is Chosen

When a treatment unambiguously defines the treatment and control groups at a given time point, its effects are usually found by comparing the two groups' mean responses. But there are many cases where the treatment timing is chosen, for which the conventional approach fails. *This paper sets up an ideal causal framework for such cases to propose a simple gamma-mixed proportional-hazard approach with three durations: the waiting time until treatment, the untreated duration from the baseline, and the treated duration from the treatment timing.* To implement the proposal, we use semiparametric piecewise-constant hazards as well as Weibull hazards with a multiplicative gamma unobserved heterogeneity affecting all three durations. Despite the three durations interwoven in complex ways, surprisingly simple closed-form likelihoods are obtained whose maximization converges well. The estimators are applied to the same data as used by Fredriksson and Johansson (2008) for employment subsidy effects on unemployment duration to find about 11.1 month reduction.

JEL Classification: C21, C41

Keywords: treatment effect, duration, treatment timing, proportional hazard, gamma heterogeneity

Corresponding author:

Per Johansson
IFAU
Box 513
75120 Uppsala
Sweden
E-mail: per.johansson@ifau.uu.se

1 Introduction

In the treatment effect literature (see Rosenbaum 2002, Lee 2005, Pearl 2009 and Imbens and Wooldridge 2009, among many others), typically there is a treatment unambiguously defining the treatment and control groups at a given time point, and its effect on a response variable is usually found by comparing the two groups' mean responses. But there are many cases where the treatment timing is chosen (by the individuals), and the conventional approach fails in such cases.

For instance, in a job training for the unemployed, the unemployed search for new jobs initially, and then enrol for the training later; here, when to take the training is a choice variable. The conventional approach (motivated by statistical experiments where a treatment is administered at a fixed time) would classify the unemployed as a treated or a control, depending on whether they enrol for the program or not over a time interval determined by the researcher. But choosing the time interval is inherently arbitrary, because the more unemployed persons fall in the treatment group as the interval gets larger. Another example is the effect of having a child on the marital duration (Lillard 1993), where the timing of the treatment (having a child) matters much for the marital duration. Similar problems appear in medical treatments; e.g., a medical treatment is taken by the diseased, but when to take it is up to the diseased or the treatment administrator. Ignoring the treatment timing issue in these cases will lead to false inferences.

When treatment timing is chosen, there appears a waiting duration W until treatment. With the response variable of interest being a survival duration, there arise three durations: W , the untreated response duration Y^{o*} starting at the baseline, and the treated response duration Y^{w*} starting after the treatment at $W = w$; there are many potential treated durations Y^{w*} for $w \in (0, \infty)$. With W in, the usual "static" definition of treatment effect, say $Y^{treated} - Y^{untreated}$, does not work because the untreated (i.e., never treated) group ($W > Y^{o*}$) has a relatively short Y^{o*} . Even if we use the group not treated until a fixed time $w > 0$, there still occurs the same problem of short Y^{o*} due to $w > Y^{o*}$.

The goal of this paper is two-fold. First, we set up an “ideal” causal framework for the complex cases involving three durations. Our approach is ideal in the sense that we compare Y^{o*} to the ‘counter-factual Y^{w*} with $w = 0$ ’ as if there were no waiting, which removes the complicating treatment timing out of the picture altogether. Second, to implement the causal framework, we propose a ‘gamma-mixed proportional-hazard’ approach. The gamma-distributed unobserved factor leads to a closed-form likelihood function despite the three durations interwoven in complex ways so that maximum likelihood estimator (MLE) can be applied straightforwardly. With hazard functions specified as a piecewise-constant baseline hazard multiplied by an exponential function, the only substantial assumption in the MLE is the gamma assumption.

The rest of this paper is organized as follows. Section 2 provides the details of our causal approach, which is then compared with related studies in the literature. Section 3 sets up our basic framework and notation. Section 4 presents the general forms of the likelihood components with no unobserved heterogeneity, and then those under Weibull and piecewise-constant hazards. Section 5 introduces a gamma unobserved heterogeneity to obtain the aforementioned closed-form likelihoods. Section 6 presents our empirical analysis. Finally, Section 7 concludes. Lengthy derivations can be found in the appendix.

2 Causality at Baseline and Related Studies

When the treatment timing W is chosen, our “causality at-baseline” is using

$$Y^{+*} - Y^{o*}$$

as the basis for treatment effects, where Y^{+} is the treated duration with no (“or 0^+ ”) waiting; the smallest value for W in our data is one day that amounts to 0^+ . That is, $Y^{+*} - Y^{o*}$ is the difference between the treated and untreated durations starting at the baseline.*

This causal framework is ideal because it removes the complicating treatment timing aspect as if the treatment were taken at the baseline to compare Y^{o*} to Y^{+*} ;

essentially, this restores the static $Y^{treated} - Y^{untreated}$. For a medical treatment and a patient regretting not having taken the medicine sooner, this answers the question “what if I had taken the medicine immediately”. The control group at the baseline (i.e., the entire study population, as nobody gets treated at the baseline) is opposite to the never-treated group ($W > Y^{o*}$) that is the control group found only at the end of the study; the latter has the aforementioned problem of a relatively short Y^{o*} that the former does not have.

To implement the approach, we first adopt a Weibull proportional hazard model with a gamma-distributed unobserved heterogeneity (a “factor”). Then we relax the Weibull specification with a semiparametric piecewise-constant hazard. So long as the gamma factor is maintained, both specifications lead to closed-form likelihood functions whose maximization converges well. The proposed method is then applied to the same data set as used in Fredriksson and Johansson (2008) and De Luna and Johansson (2010).

Having closed-form likelihoods saves the computation time enormously when the data at hand is as huge as ours: 631,358 observations covering the entire Swedish population. Without closed-form likelihoods, one can use a simulated MLE, which however takes many times longer, resulting in a difference of several weeks in computation time, if not months. With three durations entangled, it is unlikely to find closed-form likelihoods without gamma distribution in a mixed proportional hazard set-up. *Thus our estimators provide simple—perhaps the simplest possible—approaches to the complicated treatment-timing issue with three durations.*

Differently from our approach, Fredriksson and Johansson (2008) proposed a non-parametric matching estimator, whose distribution was found by De Luna and Johansson (2010): the treated and the untreated are compared on a given $W = w$, resulting in the mean effect on the treated for the given w via hazard and survival functions. Since $P(W = 0) = 0$, the nonparametric approach in Fredriksson and Johansson (2008) could not find Y^{+*} , whereas we parametrize the dependence of Y^{w*} on w to find Y^{+*} by “interpolating Y^{w*} ” back to $w = 0$. Although such a parametrization is restrictive, it makes the causal analysis possible with $Y^{+*} - Y^{o*}$. Also Fredriksson and Johansson

(2008) could not allow unobserved factors.

Eberwein et al. (1997) modeled treatment-timing choice in discrete time: each period’s survival is modeled as logit, and the total survival probability is the logit products. Eberwein et al. (1997) used discrete unobserved factors with two or three support points, but our experience with discrete factors was not good with convergence problems—the literature is fairly mute on this problem though. In view of the fact that only a slope parameter can be identified for an observed covariate, the root cause of this convergence problem with discrete factors would be trying to identify too many parameters associated with them.

Abbring and Van den Berg (2003b) examined nonparametric identification (but not estimation) of treatment effect when the treatment timing is chosen: they adopted mixed proportional hazard models in continuous time to define the treatment effect as a ‘hazard shifter’. In contrast to our causal approach with *three* durations, Abbring and Van den Berg (2003b) used only *two* durations: W and the potential response duration when treated at $W = w$. Two durations are not enough to lead to our $Y^{+*} - Y^{o*}$, which might be why Abbring and Van den Berg (2003b) defined the treatment effect as a hazard shifter, not $Y^{+*} - Y^{o*}$. But the following close examination of their two-stage identification reveals that there appear, in fact, three durations.

The identification in Abbring and Van den Berg (2003b) is done in two stages. In the first stage (Proposition 2 in p.1506), they identify the *untreated* hazard using the competing risk framework in Abbring and Van den Berg (2003a) where $\min(W, Y^{o*})$ is observed. In the second stage, the *treated* hazard is modelled (p.1503) as ‘the untreated hazard found in the first-stage’ times a ‘multiplicative hazard shifter’ to capture the shift in the hazard due to the treatment at $W = w$; the second-stage identifies this multiplicative shifter. Hence only the untreated duration appears in the first stage and only the treated duration appears in the second stage to result in three durations in total, although the distinction between Y^{o*} and Y^{w*} is not explicitly made in Abbring and Van den Berg (2003b).

The lack of distinction between Y^{o*} and Y^{w*} seems to have caused some problems later on, when the identification result of Abbring and Van den Berg (2003b) was

implemented in a single-stage nonparametric MLE with simulated data by Gaure et al. (2007): they used only models under no treatment effect (i.e., $Y^{o*} = Y^{w*}$). We suspect that the single-stage nonparametric MLE would not work if the treatment effect is not zero (i.e., if $Y^{o*} \neq Y^{w*}$). Intuitively, by using a single stage instead of two stages, some “mix-up” would occur when $W < Y^{o*}$ so that the treated duration Y^{w*} is observed along with W ; in contrast, in the first competing-risk stage of Abbring and Van den Berg (2003b), only W (but not Y^{w*}) is used if $W < Y^{o*}$. When we implement our MLE later, we will also adopt a two-stage procedure using only W and Y^{o*} first and then (W, Y^{o*}, Y^{w*}) later. The single-stage MLE did not work well in our experience, which seems closely related to the fact that the nonparametric identification of Abbring and Van den Berg (2003b) was done also in analogous two stages.

Before proceeding further, we note three shortcomings of our method. The first is not allowing time-varying covariates; although some modifications of our MLE may allow time-varying covariates, such modifications will no longer yield a simple easy-to-use MLE. The second is that no future treatment should affect the past; if it did, we would have Y^{w*} starting well before w . The third is that the treatment should be administered instantly at w (to stay on); if the treatment itself takes a long time to administer, then we would need yet another duration—four in total.

3 Durations and Random Right Censoring

3.1 Notation and Underlying Durations

With X being time-constant covariates including 1 as its first element, suppose that the waiting time $W_i|X_i$ of person i until treatment has a distribution function (df) $P(W \leq w|X_i) = F_{W|X_i}(w; \alpha)$ where α is parameters. The untreated duration $Y_i^{o*}|X_i$ has a df $P(Y^{o*} \leq y|X_i) = F_{Y^{o*}|X_i}(y; \beta)$ where β is parameters. For the control group ($W_i > Y_i^{o*}$), Y_i^{o*} is observed but W_i is not. For the treatment group ($W_i \leq Y_i^{o*}$), W_i is observed but Y_i^{o*} is not, and the duration starting from $W_i = w$ (i.e., the ‘*net treated duration*’) $Y_i^{w*}|X_i$ has a df $P(Y^{w*} \leq y|X_i) = F_{Y^{w*}|X_i}(y; w, \tau)$ where τ is parameters.

Assume that the distribution functions are twice continuously differentiable for the usual MLE regularity conditions.

It would be more precise to write, e.g., $P(W \leq w | X_i = x) = F_{W|X}(w|x; \alpha)$ to make $X_i = x$ explicit. But, since X_i will be always conditioned on at its observed value, we will simply write $F_{W|X}(w; \alpha)$ to omit “ $|x$ ” before α unless this causes confusion. Often “ $|\alpha$ ”, “ $|\beta$ ” and “ $|\tau$ ” will be omitted as well to simplify notation. In the following, the subscript i will be suppressed unless otherwise necessary, as ‘*iid* (independent and identically distributed)’ will be maintained throughout. Also, df’s, survival functions (i.e., one minus the distribution function) and densities will be denoted by F , S and f , respectively, with the appropriate subscripts.

Whereas the net treated duration starting from $W = w$ is Y^{w*} , the gross treated duration starting from 0 and its “random version” are

$$w + Y^{w*} \quad \text{and} \quad W + Y^{W*}.$$

We will assume for a while that

$$W, Y^{o*}, Y^{w*} \text{ are independent of one another given } X \text{ for any value of } w. \quad (3.1)$$

This is a ‘selection-on-observable’ assumption, as W plays the role of the treatment indicator.

A restrictive feature of (3.1) is that Y^{o*} is independent of Y^{w*} given X , which will be relaxed later when Y^{o*} and Y^{w*} (as well as W) are allowed to be related by sharing an unobserved heterogeneity δ . Other than this, relaxing (3.1) would be possible only at the cost of getting non-closed-form likelihood functions that are far more complicated than those to appear below. Notice that (3.1) is not conditioned on $W = w$; e.g., $(W, Y^{o*}, Y^{1.2*})$ are independent (with $w = 1.2$) of one another as well as $(W, Y^{o*}, Y^{7.7*})$ are so (with $w = 7.7$).

3.2 Random Right Censoring and Observed Durations

Typically, there exists a right censoring problem in duration data, which requires modifying the preceding basic setup. Consider a censoring duration C from $F_{C|X}$ where

no α , β and τ parameters appear. With ‘ $X \perp\!\!\!\perp Y|Z$ ’ denoting the independence between X and Y given Z , assume the usual non-informative censoring: for any value of w ,

$$C \perp\!\!\!\perp (W, Y^{o*}, Y^{w*})|X. \quad (3.2)$$

With $(W, Y^{o*}, Y^{w*})|X$ being independent of one another, *our main assumption combining the selection-on-observables (3.1) and the non-informative censoring (3.2) is*

$$C, W, Y^{o*}, Y^{w*} \text{ are independent of one another given } X \text{ for any value of } w. \quad (A_o)$$

No relation ‘ $Y^{o*} \perp\!\!\!\perp Y^{w*}|X$ ’ in A_o needs some discussion as in the next paragraph, because it is opposite to the perfect relation $Y^{o*} = Y^{w*}$ or its intercept-shifted version $Y^{w*} = Y^{o*} + \omega w$ for a constant ω that are often used. Note that the relationship between Y^{o*} and Y^{w*} cannot be nonparametrically identified as only one of them is observed for a given individual, which is why the two polar (perfect and zero relation) cases are considered.

The strongest definition of no treatment effect is $Y^{o*} = Y^{w*}$, and a weaker version is the exchangeability of Y^{o*} and Y^{w*} as exchangeability means “indistinguishability” between Y^{o*} and Y^{w*} ; exchangeability implies the same marginal distribution for Y^{o*} and Y^{w*} , but it is mute on the relation between Y^{o*} and Y^{w*} . Under perfect relation, no effect is $Y^{o*} = Y^{w*}$, while some effect can be modeled by specifying Y^{w*} as a deterministic function of Y^{o*} and w as in $Y^{w*} = Y^{o*} + \omega w$. Under $Y^{o*} \perp\!\!\!\perp Y^{w*}|X$, no effect is the same marginal distribution of Y^{o*} and Y^{w*} (as this implies the exchangeability of Y^{o*} and Y^{w*}), while some effect can be modeled by any distribution of Y^{w*} different from that of Y^{o*} . If we proceed under perfect relation, our task becomes simpler as one fewer duration appears. But if W is long, $Y^{o*} = Y^{w*}$ would be implausible as “the world has changed” after the long W , whereas $Y^{o*} \perp\!\!\!\perp Y^{w*}|X$ may become more plausible as, indeed, the world has changed.

Under right random censoring, define the observed untreated duration Y^0 and the observed gross treated duration Y^1 :

$$Y^0 \equiv \min(Y^{o*}, C) \quad \text{and} \quad Y^1 \equiv \min(W + Y^{w*}, C).$$

Let $1[A] = 1$ if A holds and 0 otherwise. Then the treatment dummy is

$$D \equiv 1[W \leq Y^0] = 1[W \leq \min(Y^{o*}, C)].$$

Note $D \neq 1[W \leq Y^{o*}]$ because a control individual may have $W \leq Y^{o*}$ as well: a censored control ($C < \min(Y^{o*}, W)$) may get treated between C and Y^{o*} if $C < W < Y^{o*}$, but this will never be known and what is observed is only that the individual is untreated up to C .

Define the non-censoring dummy variables

$$Q_0 \equiv 1[Y^{o*} \leq C], \quad Q_1 \equiv 1[W + Y^{w*} \leq C] \quad \text{and} \quad Q \equiv (1 - D)Q_0 + DQ_1.$$

Finally, define the observed response

$$\begin{aligned} Y &= (1 - D)Y^0 + DY^1 \\ &= 1[\min(Y^{o*}, C) < W] \cdot \min(Y^{o*}, C) + 1[W \leq \min(Y^{o*}, C)] \cdot \min(W + Y^{w*}, C). \end{aligned}$$

4 Likelihood Functions for Four Groups

4.1 General Likelihood Functions

Under random right censoring, there are four groups, depending on the treatment status $D = 0, 1$ and non-censoring indicator $Q = 0, 1$ (the following four diagrams help understanding what is observed in each group). We present the X -conditional likelihood function for each group, whose detailed derivation can be found in the appendix.

$D = 0, Q = 0$	$D = 0, Q = 1$	$D = 1, Q = 0$	$D = 1, Q = 1$
$W \longrightarrow \longrightarrow$	$W \longrightarrow \dashrightarrow$	$W \longrightarrow$	$W \longrightarrow$
$Y^{o*} \longrightarrow \longrightarrow$	$Y^{o*} \longrightarrow$	$Y^{o*} \longrightarrow \dashrightarrow$	$Y^{o*} \longrightarrow \dashrightarrow$
Y^{w*}	Y^{w*}	$Y^{w*} \longrightarrow \longrightarrow$	$Y^{w*} \longrightarrow$
$C \longrightarrow$	$C \longrightarrow \longrightarrow$	$C \longrightarrow \longrightarrow$	$C \longrightarrow \longrightarrow \longrightarrow$
$C = y$ observed	$Y^{o*} = y$ obs	$W = w, C = y$ obs	$W = w, w + Y^{w*} = y$

For the *censored control group*, the observed duration is the censoring duration y , and

$$D = 0, Q = 0 \iff \min(Y^{o*}, C) < W, C < Y^{o*} \iff C < W, C < Y^{o*}.$$

Recalling the independence among $(W, Y^{o*}, C)|X$, the likelihood function for $C = y$ is

$$S_{W|X}(y)S_{Y^{o*}|X}(y) \cdot f_{C|X}(y). \quad (4.1)$$

For the *uncensored control group*, the observed duration y is the untreated duration, and

$$D = 0, Q = 1 \iff \min(Y^{o*}, C) < W, Y^{o*} < C \iff Y^{o*} < W, Y^{o*} < C.$$

The likelihood function for $Y^{o*} = y$ is

$$S_{W|X}(y)S_{C|X}(y) \cdot f_{Y^{o*}|X}(y). \quad (4.2)$$

For the *censored treated group*, the observed durations are w and the censoring duration y , and

$$D = 1, Q = 0 \iff W < \min(Y^{o*}, C), C < W + Y^{W*} \iff W < \min(Y^{o*}, C), C - W < Y^{W*}.$$

Since ' $W = w \geq C = y$ ' cannot happen to the treatment group, the likelihood function for $W = w$ and $C = y$ with $w < y$ is

$$S_{Y^{o*}|X}(w)S_{Y^{w*}|X}(y - w)1[w < y] \cdot f_{W|X}(w)f_{C|X}(y). \quad (4.3)$$

For the *uncensored treated group*, the observed durations are w and the gross treated duration y , and

$$D = 1, Q = 1 \iff W < \min(Y^{o*}, C), W + Y^{W*} < C \iff W < Y^{o*}, W + Y^{W*} < C.$$

Since ' $W = w \geq W + Y^{W*} = y$ ' cannot happen to the uncensored treated group, the likelihood function for $(W = w, W + Y^{W*} = y) \iff (W = w, Y^{w*} = y - w)$ is

$$S_{Y^{o*}|X}(w)S_{C|X}(y)1[w < y] \cdot f_{W|X}(w)f_{Y^{w*}|X}(y - w). \quad (4.4)$$

4.2 Likelihood Function with Weibull Hazards

For a Weibull specification for W , assume

$$W|X \sim Weibull\{\alpha_0, \alpha(X)\} \quad \text{where } \alpha_0 > 0 \text{ and } \alpha(X) = \exp(X'\alpha_x);$$

hence $S_{W|X}(w) = \exp\{-w^{\alpha_0} \exp(X'\alpha_x)\}$ and

$$f_{W|X}(w) = \alpha_0 w^{\alpha_0-1} \exp(X'\alpha_x) \cdot S_{W|X}(w) = \alpha_0 w^{\alpha_0-1} \exp(X'\alpha_x) \cdot \exp\{-w^{\alpha_0} \exp(X'\alpha_x)\}.$$

As for the untreated duration Y^{o*} , assume

$$Y^{o*}|X \sim Weibull\{\beta_0, \beta(X)\} \quad \text{where } \beta_0 > 0 \text{ and } \beta(X) = \exp(X'\beta_x);$$

$S_{Y^{o*}|X}(y) = \exp\{-y^{\beta_0} \exp(X'\beta_x)\}$ and thus

$$f_{Y^{o*}|X}(y) = \beta_0 y^{\beta_0-1} \exp(X'\beta_x) \cdot S_{Y^{o*}|X}(y) = \beta_0 y^{\beta_0-1} \exp(X'\beta_x) \cdot \exp\{-y^{\beta_0} \exp(X'\beta_x)\}.$$

For Y^{w*} , with a random vector H consisting of functions of (X, W) , assume

$$Y^{w*}|X = x \sim Weibull\{\tau_0, \tau(x, w)\}, \quad \tau_0 > 0, \tau(x, w) = \exp(h'\tau_h) \text{ with } \tau(x, 0) = \exp(x'\tau_x).$$

Although we used X instead of ' $X = x$ ' for W and Y^{o*} to simplify presentation, we use ' $X = x$ ' for Y^{w*} to prevent confusion because $Y^{w*} \neq Y^{W*}$; using w and X would not work for H , notation-wise. With a random vector M being the components of X interacting with W , an example of $h'\tau_h$ is

$$h'\tau_h = x'\tau_x + m'w\tau_{mw} + \tau_{w1}w + \tau_{w2}w^2 + \tau_{w3}w^3.$$

We have $S_{Y^{w*}|X=x}(y) = \exp\{-y^{\tau_0} \exp(h'\tau_h)\}$ and

$$f_{Y^{w*}|X=x}(y) = \tau_0 y^{\tau_0-1} \exp(h'\tau_h) \cdot S_{Y^{w*}|X=x}(y) = \tau_0 y^{\tau_0-1} \exp(h'\tau_h) \cdot \exp\{-y^{\tau_0} \exp(h'\tau_h)\}.$$

An important point for Y^{w} is that $w = 0$ gives, as all terms with w drop out of $\tau(x, w)$,*

$$Y^{+*}|X = x \sim Weibull\{\tau_0, \tau(x, 0)\} \quad \text{where } \tau(x, 0) = \exp(x'\tau_x).$$

This is the (counter-factual) treated duration when treated at the baseline. Notice that τ_0 may be specified to depend on w , although we did not allow this for simplification. If the treatment is a drug for an illness, the waiting time before taking the drug may affect the treated duration negatively or positively; if $\tau(x, 0) = \tau(x, w) \forall w$, then waiting does not affect the treated duration at all.

Putting the four likelihood components under Weibull with the censored duration parts dropped under the ignorability condition (3.2), we have

$$\begin{aligned}
D = 0, Q = 0 : S_{W|X}(y) \cdot S_{Y^{o*}|X}(y) &= \exp\{-y^{\alpha_0} \exp(X' \alpha_x)\} \cdot \exp\{-y^{\beta_0} \exp(X' \beta_x)\}; \\
D = 0, Q = 1 : S_{W|X}(y) \cdot f_{Y^{o*}|X}(y) &= \exp\{-y^{\alpha_0} \exp(X' \alpha_x)\} \\
&\quad \beta_0 y^{\beta_0 - 1} \exp(X' \beta_x) \exp\{-y^{\beta_0} \exp(X' \beta_x)\}; \\
D = 1, Q = 0 : S_{Y^{o*}|X=x}(w) S_{Y^{w*}|X=x}(y-w) \cdot 1[w < y] f_{W|X=x}(w) &\quad (4.5) \\
&= \exp\{-w^{\beta_0} \exp(x' \beta_x)\} \exp\{-(y-w)^{\tau_0} \exp(h' \tau_h)\} \cdot 1[w < y] \\
&\quad \cdot \alpha_0 w^{\alpha_0 - 1} \exp(x' \alpha_x) \exp\{-w^{\alpha_0} \exp(x' \alpha_x)\};
\end{aligned}$$

$$\begin{aligned}
D = 1, Q = 1 : S_{Y^{o*}|X=x}(w) \cdot 1[w < y] f_{W|X=x}(w) f_{Y^{w*}|X=x}(y-w) &\quad (4.6) \\
&= \exp\{-w^{\beta_0} \exp(x' \beta_x)\} \cdot 1[w < y] \cdot \alpha_0 w^{\alpha_0 - 1} \exp(x' \alpha_x) \exp\{-w^{\alpha_0} \exp(x' \alpha_x)\} \\
&\quad \cdot \tau_0 (y-w)^{\tau_0 - 1} \exp(h' \tau_h) \cdot \exp\{-(y-w)^{\tau_0} \exp(h' \tau_h)\}.
\end{aligned}$$

Treatment effect can be assessed by comparing

$$E(Y^{+*}|X) = \Gamma(1 + \tau_0^{-1}) \{\exp(X' \tau_x)\}^{-1/\tau_0} \quad \text{and} \quad E(Y^{o*}|X) = \Gamma(1 + \beta_0^{-1}) \{\exp(X' \beta_x)\}^{-1/\beta_0}$$

where $\Gamma(\eta) \equiv \int_0^\infty z^{\eta-1} e^{-z} dz$ for $\eta > 0$ is the Gamma function.

To implement this, first, obtain these expected durations for each X_i . Then the difference between the two sample averages with X integrated out using the empirical distribution can be taken as the mean effect: with ‘ $\hat{\cdot}$ ’ denoting estimators,

$$\hat{\Delta}_w \equiv \Gamma(1 + \hat{\tau}_0^{-1}) \frac{1}{N} \sum_i \exp(-X_i' \hat{\tau}_x / \hat{\tau}_0) - \Gamma(1 + \hat{\beta}_0^{-1}) \frac{1}{N} \sum_i \exp(-X_i' \hat{\beta}_x / \hat{\beta}_0).$$

4.3 Likelihood Function with Piecewise-Constant Hazards

To relax the Weibull baseline hazard assumption $\alpha_0 w^{\alpha_0-1}$ for W , consider a piecewise-constant baseline hazard $\lambda(w; \alpha_{1J})$: for a specified J and boundary points $\zeta_1 < \dots < \zeta_J$, and unknown parameters $\alpha_{1J} \equiv (\alpha_1, \alpha_2, \dots, \alpha_J)'$ along with $\alpha_0 = 1$,

$$\lambda(w; \alpha_{1J}) \equiv \alpha_0 1[w \leq \zeta_1] + \sum_{j=2}^J \alpha_{j-1} 1[\zeta_{j-1} < w \leq \zeta_j] + \alpha_J w^{\alpha_J-1} 1[\zeta_J < w]$$

where all components of α_{1J} are positive; for this, $\exp(\alpha_j)$ may be used instead of α_j . There are $J + 1$ time segments with the specified ζ_j 's, and J hazards to estimate: $\alpha_0 = 1$ is a normalization as α_0 is absorbed by the intercept in $X'\alpha_x$. The last hazard is not a constant but of Weibull form, which is to avoid the constant hazard in the last open-ended segment going to ∞ .

Integrate $\lambda(w; \alpha_{1J})$ to obtain the integrated baseline hazard: with $\zeta_0 = 0$,

$$\begin{aligned} \Lambda(w, \alpha_{1J}) &\equiv w 1[w \leq \zeta_1] + \sum_{j=2}^J \left\{ \sum_{j'=1}^{j-1} \alpha_{j'-1} (\zeta_{j'} - \zeta_{j'-1}) + \alpha_{j-1} (w - \zeta_{j-1}) \right\} \cdot 1[\zeta_{j-1} < w \leq \zeta_j] \\ &\quad + \left\{ \sum_{j'=1}^J \alpha_{j'-1} (\zeta_{j'} - \zeta_{j'-1}) + (w^{\alpha_J} - \zeta_J^{\alpha_J}) \right\} \cdot 1[\zeta_J < w] \\ &= w \quad \text{if } w \leq \zeta_1 \text{ (} w \text{ in the first segment)} \\ &= \zeta_1 + \alpha_1 (w - \zeta_1) \quad \text{if } \zeta_1 < w \leq \zeta_2 \text{ (} w \text{ in the second segment with } j = 2) \\ &\quad \dots\dots\dots \\ &= \zeta_1 + \alpha_1 (\zeta_2 - \zeta_1) + \dots + \alpha_{J-2} (\zeta_{J-1} - \zeta_{J-2}) + \alpha_{J-1} (w - \zeta_{J-1}) \quad \text{if } \zeta_{J-1} < w \leq \zeta_J \\ &= \zeta_1 + \alpha_1 (\zeta_2 - \zeta_1) + \dots + \alpha_{J-2} (\zeta_{J-1} - \zeta_{J-2}) + \alpha_{J-1} (\zeta_J - \zeta_{J-1}) + (w^{\alpha_J} - \zeta_J^{\alpha_J}) \quad \text{if } \zeta_J < w. \end{aligned}$$

The hazard function with X is then $\lambda(w; \alpha_{1J}) \exp(X'\alpha_x)$, and

$$S_{W|X}(w) = \exp\{-\Lambda(w; \alpha_{1J}) \exp(X'\alpha_x)\} \quad \text{and} \quad f_{W|X}(w) = \lambda(w; \alpha_{1J}) \exp(X'\alpha_x) \cdot S_{W|X}(w).$$

Doing analogously for Y^{o*} and Y^{w*} , the baseline hazards are $\lambda(y; \beta_{1J})$ and $\lambda(y; \tau_{1J})$. A more general model would allow different boundary points for the three durations, say, $(J_\alpha, \zeta_{\alpha j})$ for W , $(J_\beta, \zeta_{\beta j})$ for Y^{o*} and $(J_\tau, \zeta_{\tau j})$ for Y^{w*} , but we use the common

known boundary points ζ_j 's for all three durations for simplification. The four likelihood components are the same as in the Weibull case except

replacing $\gamma_0 t^{\gamma_0-1}$ and t^{γ_0} with $\lambda(t; \alpha_{1J})$ and $\Lambda(t; \alpha_{1J})$, respectively,

where γ_0 can be any of α_0 , β_0 or τ_0 , and t can be any of w , y or $y - w$.

Once all parameters are estimated, the treatment effect can be assessed using the average durations. But differently from the Weibull case, no simple formulas exist for $E(Y^{+*}|X)$ and $E(Y^{o*}|X)$. Instead, we use

$$\begin{aligned} E(Y^{+*}|X) &= \int_0^\infty S_{Y^{+*}|X}(y) dy = \int_0^\infty \exp\{-\Lambda(y; \tau_{1J}) \exp(X' \tau_x)\} dy; \\ E(Y^{o*}|X) &= \int_0^\infty S_{Y^{o*}|X}(y) dy = \int_0^\infty \exp\{-\Lambda(y; \beta_{1J}) \exp(X' \beta_x)\} dy. \end{aligned}$$

The mean effect for an individual with X is

$$E(Y^{+*}|X) - E(Y^{o*}|X) = \int_0^\infty S_{Y^{+*}|X}(y) dy - \int_0^\infty S_{Y^{o*}|X}(y) dy$$

which is the difference between the area below $S_{Y^{+*}|X}$ and the area below $S_{Y^{o*}|X}$ over $(0, \infty)$. The mean effect with X integrated out by its empirical distribution is then

$$\Delta_p \equiv \frac{1}{N} \sum_i \int_0^\infty [\exp\{-\Lambda(y; \tau_{1J}) \exp(X_i' \tau_x)\} - \exp\{-\Lambda(y; \beta_{1J}) \exp(X_i' \beta_x)\}] dy. \quad (4.7)$$

To find an estimator for Δ_p , observe that integration by parts gives, for a continuously differentiable survival function $S(y)$,

$$-\int_0^\varpi y f(y) dy = y S(y) \Big|_0^\varpi - \int_0^\varpi S(y) dy \implies \int_0^\varpi y f(y) dy = \int_0^\varpi S(y) dy \quad \text{if } \varpi S(\varpi) = 0.$$

Hence, so long as $\varpi S_{Y^{o*}|X}(\varpi) = 0 = \varpi S_{Y^{+*}|X}(\varpi)$ for some maximum duration ϖ , we can draw uniform random numbers over $(0, \varpi)$ to use a Monte-Carlo-integration estimator for Δ_p :

$$\hat{\Delta}_p \equiv \frac{1}{N} \sum_{i=1}^N \frac{1}{n} \sum_{\nu=1}^n [\exp\{-\Lambda(u_\nu; \hat{\tau}_{1J}) \exp(X_i' \hat{\tau}_x)\} - \exp\{-\Lambda(u_\nu; \hat{\beta}_{1J}) \exp(X_i' \hat{\beta}_x)\}] \quad (4.8)$$

where u_ν , $\nu = 1, \dots, n$, are uniform random numbers over $(0, \varpi)$.

5 Gamma Unobserved Heterogeneity

The main restriction so far has been the independence (3.1) of (W, Y^{o*}, Y^{w*}) given X . To relax this, this section introduces an unobserved factor δ through which the three durations are allowed to be related. Eberwein et al. (1997) did essentially the same task, but in a discrete time framework with a discrete unobserved factor. Estimating the support points of a discrete factor is, however, problematic for the MLE asymptotic distribution. For instance, with two support points $s_1 < s_2$, the parameter space for s_1 becomes $(-\infty, s_2)$ whose upper bound depends on s_2 ; this goes against the usual regularity conditions of MLE.

In the following, firstly, useful facts about gamma distribution and integrations involving exponential functions of unit-mean gamma are provided to be used later. Secondly, δ is introduced into the Weibull model for only (W, Y^{o*}) to show that α and β are identified regardless of Y^{w*} . This is necessary for estimating α and β first and then τ later; this two-stage estimation is reminiscent of the two-stage identification of Abbring and Van den Berg (2003b) as was already mentioned. Thirdly, δ is considered for all of (W, Y^{o*}, Y^{w*}) under the Weibull hazard. Fourthly, δ in (W, Y^{o*}, Y^{w*}) is examined for the piecewise-constant hazards.

5.1 Integration Involving Unit-Mean Gamma Distribution

In gamma distribution denoted $Gamma(\eta, \theta)$, $\eta > 0$ is the ‘shape parameter’ and $\theta > 0$ is the ‘scale parameter’. With $Z \sim Gamma(\eta, \theta)$, the density is

$$f(z) = \frac{1}{\Gamma(\eta)\theta} \left(\frac{z}{\theta}\right)^{\eta-1} e^{-(z/\theta)} \quad (\Gamma(\eta) \text{ is the normalizing factor})$$

$$\implies E(Z) = \eta\theta, \quad V(Z) = \eta\theta^2 \quad \text{and} \quad \mu Z \sim Gamma(\eta, \mu\theta) \text{ for any positive constant } \mu.$$

If $\theta^{-1} = \eta$ in $Z \sim Gamma(\eta, \theta)$, then $E(Z) = 1$ and $V(Z) = \eta^{-1}$ with the density

$$f(z) = \frac{\eta^\eta}{\Gamma(\eta)} z^{\eta-1} e^{-\eta z}.$$

$Gamma(\eta, \eta^{-1})$ will appear for a multiplicative unobserved heterogeneity because $E(Z) = 1$ is needed for a normalization.

Three useful facts about $Z \sim \text{Gamma}(\eta, \eta^{-1})$ are that, for any constant p ,

$$E\{\exp(-pZ)\} = (\eta^{-1}p + 1)^{-\eta}; \quad (5.1)$$

$$E\{Z \exp(-pZ)\} = (\eta^{-1}p + 1)^{-\eta-1}; \quad (5.2)$$

$$E\{Z^2 \exp(-pZ)\} = (1 + \eta^{-1}) \cdot (\eta^{-1}p + 1)^{-\eta-2}; \quad (5.3)$$

the proofs are in the appendix. These equations are crucial for closed-form likelihoods when a gamma factor δ gets integrated out. Gamma δ 's can be seen in Lancaster (1979), Heckman and Singer (1984) and Han and Hausman (1990). The convenience of a gamma δ was shown by Lancaster (1979), which was only for a single duration using (5.1), whereas our set-up involves three durations using all of (5.1), (5.2) and (5.3). A theoretical support for multiplicative gamma δ 's was provided by Abbring and Van den Berg (2007).

5.2 Unobserved Heterogeneity for Weibull (W, Y^{o*})

Define

$$\begin{aligned} G &\equiv (1 - D)Y + DW = (1 - D)Y^0 + DW = (1 - D) \min(Y^{o*}, C) + DW \\ &= (1 - D)(1 - Q)C + (1 - D)QY^{o*} + DW. \end{aligned} \quad (5.4)$$

Using the three cases $(1 - D)(1 - Q)$, $(1 - D)Q$ and D , the α and β parameters can be estimated without involving the τ parameters.

Suppose $\delta \sim \text{Gamma}(\eta, \eta^{-1})$ and

$$W|(X, \delta) \sim \text{Weibull}\{\alpha_0, \alpha(X, \delta)\} \quad \text{where } \alpha_0 > 0 \text{ and } \alpha(X, \delta) = \delta \exp(X'\alpha_x); \quad (5.5)$$

$$S_{W|X, \delta}(w) = \exp\{-w^{\alpha_0} \delta \exp(X'\alpha_x)\}, \quad f_{W|X, \delta}(w) = \alpha_0 w^{\alpha_0-1} \delta \exp(X'\alpha_x) \exp\{-w^{\alpha_0} \delta \exp(X'\alpha_x)\}.$$

Using $\delta \sim \text{Gamma}(\eta, \eta^{-1})$ is a normalization because, if $\bar{\delta} \exp(X'\alpha_x)$ holds for $\bar{\delta} \sim \text{Gamma}(\eta, \theta)$, then we can rewrite always

$$\bar{\delta} \exp(X'\alpha_x) = \frac{\bar{\delta}}{E(\bar{\delta})} \exp\{X'\alpha_x + \ln E(\bar{\delta})\} = \delta \exp(X'\alpha_x)$$

with $\ln E(\bar{\delta})$ absorbed into the intercept in $X'\alpha_x$.

Assume $C \Pi \delta | X$ and $W \Pi Y^{o*} | (X, \delta)$ ($(W, Y^{o*}) | X$ are allowed to be related through δ)—a stronger version of this assumption to be used mainly will appear in A_u below. Suppose

$$Y^{o*} | (X, \delta) \sim \text{Weibull}\{\beta_0, \beta(X, \delta)\} \quad \text{where } \beta_0 > 0 \text{ and } \beta(X, \delta) = \delta \exp(X' \beta_x). \quad (5.6)$$

It is more general to set $\beta(X, \delta) = \mu \delta \exp(X' \beta_x)$ for a parameter μ . But, with the normalization $\mu \delta / E(\mu \delta)$, μ drops out, i.e., $\ln E(\mu \delta)$ is absorbed by the intercept in $X' \beta_x$. Hence

$$S_{Y^{o*} | X, \delta}(y) = \exp\{-y^{\beta_0} \delta \exp(X' \beta_x)\}, \quad f_{Y^{o*} | X, \delta}(y) = \beta_0 y^{\beta_0 - 1} \delta \exp(X' \beta_x) \exp\{-y^{\beta_0} \delta \exp(X' \beta_x)\}.$$

The (X, δ) -conditioned likelihood functions for $G = g$ are

$$\begin{aligned} D = 0, Q = 0 \text{ with } C = g &: S_{W | X, \delta}(g) S_{Y^{o*} | X, \delta}(g) f_{C | X}(g); \\ D = 0, Q = 1 \text{ with } Y^{o*} = g &: S_{W | X, \delta}(g) S_{C | X}(g) f_{Y^{o*} | X, \delta}(g); \\ D = 1 \text{ with } W = g &: S_{Y^{o*} | X, \delta}(g) S_{C | X}(g) f_{W | X, \delta}(g). \end{aligned} \quad (5.7)$$

For $D = 0$ and $Q = 0$, we have from (5.5) and (5.6)

$$S_{W | X, \delta}(g) S_{Y^{o*} | X, \delta}(g) = \exp[-\{g^{\alpha_0} \exp(X' \alpha_x) + g^{\beta_0} \exp(X' \beta_x)\} \delta].$$

Its δ -free version is, using (5.1) with $p = g^{\alpha_0} \exp(X' \alpha_x) + g^{\beta_0} \exp(X' \beta_x)$,

$$E\{S_{W | X, \delta}(g) S_{Y^{o*} | X, \delta}(g) | X\} = \{\eta^{-1} g^{\alpha_0} \exp(X' \alpha_x) + \eta^{-1} g^{\beta_0} \exp(X' \beta_x) + 1\}^{-\eta}. \quad (5.8)$$

As for $D = 0$ and $Q = 1$, observe

$$S_{W | X, \delta}(g) f_{Y^{o*} | X, \delta}(g) = \beta_0 g^{\beta_0 - 1} \exp(X' \beta_x) \cdot \delta \exp[-\{g^{\alpha_0} \exp(X' \alpha_x) + g^{\beta_0} \exp(X' \beta_x)\} \delta].$$

Its δ -free version is, using (5.2) with $p = g^{\alpha_0} \exp(X' \alpha_x) + g^{\beta_0} \exp(X' \beta_x)$,

$$\begin{aligned} E\{S_{W | X, \delta}(g) f_{Y^{o*} | X, \delta}(g) | X\} \\ = \beta_0 g^{\beta_0 - 1} \exp(X' \beta_x) \cdot \{\eta^{-1} g^{\alpha_0} \exp(X' \alpha_x) + \eta^{-1} g^{\beta_0} \exp(X' \beta_x) + 1\}^{-\eta - 1}. \end{aligned} \quad (5.9)$$

For $D = 1$, observe

$$S_{Y^{o*} | X, \delta}(g) f_{W | X, \delta}(g) = \alpha_0 g^{\alpha_0 - 1} \exp(X' \alpha_x) \cdot \delta \exp[-\{g^{\alpha_0} \exp(X' \alpha_x) + g^{\beta_0} \exp(X' \beta_x)\} \delta].$$

Its δ -free version is, using (5.2) with $p = g^{\alpha_0} \exp(X' \alpha_x) + g^{\beta_0} \exp(X' \beta_x)$,

$$\begin{aligned} E\{S_{Y^{o*}|X,\delta}(g)f_{W|X,\delta}(g) | X\} \\ = \alpha_0 g^{\alpha_0-1} \exp(X' \alpha_x) \cdot \{\eta^{-1} g^{\alpha_0} \exp(X' \alpha_x) + \eta^{-1} g^{\beta_0} \exp(X' \beta_x) + 1\}^{-\eta-1}. \end{aligned}$$

5.3 Unobserved Heterogeneity for Weibull (W, Y^{o*}, Y^{w*})

With Y^{w*} considered along with (W, Y^{o*}) , assume (compare to A_o)

$$C \perp \delta | X \quad \text{and} \quad (C, W, Y^{o*}, Y^{w*}) | (X, \delta) \text{ are independent of one another for any } w \quad (A_u)$$

which is a ‘selection-on-unobservable’ assumption combined with non-informative censoring. This allows (W, Y^{o*}, Y^{w*}) to be related through δ as well as X , whereas C can be related to (W, Y^{o*}, Y^{w*}) only through X .

Suppose, making $X = x$ explicit along with $W = w$,

$$\begin{aligned} Y^{w*} | (X = x, \delta) \sim \text{Weibull}\{\tau_0, \tau(x, w, \delta)\} \quad \text{where } \tau_0 > 0 \text{ and } \tau(x, w, \delta) = \delta \exp(h' \tau_h); \\ S_{Y^{w*}|X=x,\delta}(y) = \exp\{-y^{\tau_0} \delta \exp(h' \tau_h)\}, \quad f_{Y^{w*}|X=x,\delta}(y) = \tau_0 y^{\tau_0-1} \delta \exp(h' \tau_h) \exp\{-y^{\tau_0} \delta \exp(h' \tau_h)\}. \end{aligned}$$

For the control group, the δ -free likelihoods are the same as (5.8) and (5.9) except the notation change from g to y :

$$\begin{aligned} D = 0, Q = 0 (C = y) : \{\eta^{-1} y^{\alpha_0} \exp(X' \alpha_x) + \eta^{-1} y^{\beta_0} \exp(X' \beta_x) + 1\}^{-\eta}; \\ D = 0, Q = 1 (Y^{o*} = y) : \beta_0 y^{\beta_0-1} \exp(X' \beta_x) \{\eta^{-1} y^{\alpha_0} \exp(X' \alpha_x) + \eta^{-1} y^{\beta_0} \exp(X' \beta_x) + 1\}^{-\eta-1}. \end{aligned}$$

The appendix shows that, for $(D = 1, Q = 0)$, the δ -free likelihood for $(W = w, C = y)$ is

$$\begin{aligned} 1[w < y] \alpha_0 w^{(\alpha_0-1)} \exp(x' \alpha_x) \\ \cdot \{\eta^{-1} w^{\alpha_0} \exp(x' \alpha_x) + \eta^{-1} w^{\beta_0} \exp(x' \beta_x) + \eta^{-1} (y - w)^{\tau_0} \exp(h' \tau_h) + 1\}^{-\eta-1} \end{aligned} \quad (5.10)$$

and that, for $(D = 1, Q = 1)$, the δ -free likelihood for $(W = w, W + Y^{W*} = y)$ is

$$\begin{aligned} \alpha_0 w^{(\alpha_0-1)} \exp(x' \alpha_x) \cdot \tau_0 (y - w)^{\tau_0-1} \exp(h' \tau_h) \cdot 1[w < y] \\ \cdot (1 + \eta^{-1}) \{\eta^{-1} w^{\alpha_0} \exp(x' \alpha_x) + \eta^{-1} w^{\beta_0} \exp(x' \beta_x) + \eta^{-1} (y - w)^{\tau_0} \exp(h' \tau_h) + 1\}^{-\eta-2}. \end{aligned} \quad (5.11)$$

Turning to treatment effects, we present two options. One comes from integrating out δ :

$$\begin{aligned} E(Y^{+*}|X) &= E\{ E(Y^{+*}|\delta, X) |X\} = \int_0^\infty \int_0^\infty S_{Y^{+*}|\delta, X}(y) dy \frac{\eta^\eta}{\Gamma(\eta)} \delta^{\eta-1} e^{-\eta\delta} d\delta \\ &= \int_0^\infty E[\exp\{-y^{\tau_0} \delta \exp(X'\tau_x)\} |X] dy = \int_0^\infty \{\eta^{-1} y^{\tau_0} \exp(X'\tau_x) + 1\}^{-\eta} dy; \\ E(Y^{o*}|X) &= E\{ E(Y^{o*}|\delta, X) |X\} = \int_0^\infty \{\eta^{-1} y^{\beta_0} \exp(X'\beta_x) + 1\}^{-\eta} dy. \end{aligned}$$

The mean effect $\Delta_{w\delta}$ with X integrated out by its empirical distribution and an estimator $\hat{\Delta}_{w\delta}$ using $u_\nu \sim U[0, \varpi]$, $\nu = 1, \dots, n$, are (recall (4.7) and (4.8))

$$\begin{aligned} \Delta_{w\delta} &\equiv \frac{1}{N} \sum_i \int_0^\infty \{\eta^{-1} y^{\tau_0} \exp(X'_i \tau_x) + 1\}^{-\eta} dy - \frac{1}{N} \sum_i \{\eta^{-1} y^{\beta_0} \exp(X'_i \beta_x) + 1\}^{-\eta} dy, \\ \hat{\Delta}_{w\delta} &\equiv \frac{1}{N} \sum_{i=1}^N \frac{1}{n} \sum_{\nu=1}^n [\{\hat{\eta}^{-1} u_\nu^{\hat{\tau}_0} \exp(X'_i \hat{\tau}_x) + 1\}^{-\hat{\eta}} - \{\hat{\eta}^{-1} u_\nu^{\hat{\beta}_0} \exp(X'_i \hat{\beta}_x) + 1\}^{-\hat{\eta}}]. \end{aligned} \quad (5.12)$$

The other is based on the *proportional hazard-change relative to the untreated hazard*:

$$\Delta_{wh}(t) \equiv \frac{\tau_0 t^{\tau_0-1} \delta \exp(X'\tau_x) - \beta_0 t^{\beta_0-1} \delta \exp(X'\beta_x)}{\beta_0 t^{\beta_0-1} \delta \exp(X'\beta_x)} = \frac{\tau_0}{\beta_0} t^{\tau_0-\beta_0} \exp\{X'(\tau_x - \beta_x)\} - 1$$

that is particularly attractive as it is free of δ . The X -integrated sample version is

$$\hat{\Delta}_{wh}(t) \equiv \frac{\hat{\tau}_0}{\hat{\beta}_0} t^{\hat{\tau}_0-\hat{\beta}_0} \frac{1}{N} \sum_i \exp\{X'_i(\hat{\tau}_x - \hat{\beta}_x)\} - 1; \quad (5.13)$$

since this is a function of time t , it would be informative to plot this against t . As $\Delta_{wh}(t)$ shows the hazard shift due to the treatment, $\Delta_{wh}(t)$ is essentially the same as the treatment effect definition in Abbring and Van den Berg (2003b).

5.4 Unobserved Heterogeneity with Piecewise-Constant Hazards

For the piecewise-constant hazard with $\delta \sim \text{Gamma}(\eta, \eta^{-1})$, we just need to replace the Weibull baseline and integrated hazards with λ and Λ , respectively, as follows to obtain the δ -free likelihoods:

$$D = 0, Q = 0 : \{\eta^{-1} \Lambda(y; \alpha_{1J}) \exp(X'\alpha_x) + \eta^{-1} \Lambda(y; \beta_{1J}) \exp(X'\beta_x) + 1\}^{-\eta}; \quad (5.14)$$

$$D = 0, Q = 1 : \lambda(y; \beta_{1J}) \exp(X'\beta_x) \cdot \{\eta^{-1} \Lambda(t; \alpha_{1J}) \exp(X'\alpha_x) + \eta^{-1} \Lambda(y; \beta_{1J}) \exp(X'\beta_x) + 1\}^{-\eta-1}.$$

For the censored treated ($D = 1, Q = 0$),

$$1[w < y]\lambda(w; \alpha_{1J}) \exp(x'\alpha_x) \cdot \{\eta^{-1}\Lambda(w; \alpha_{1J}) \exp(x'\alpha_x) + \eta^{-1}\Lambda(w; \beta_{1J}) \exp(x'\beta_x) + \eta^{-1}\Lambda(y - w; \tau_{1J}) \exp(h'\tau_h) + 1\}^{-\eta-1}. \quad (5.15)$$

For the uncensored treated ($D = 1, Q = 1$),

$$\lambda(w; \alpha_{1J}) \exp(x'\alpha_x) \cdot \lambda(y - w; \tau_{1J}) \exp(h'\tau_h) \cdot 1[w < y] \cdot (1 + \eta^{-1})\{\eta^{-1}\Lambda(w; \alpha_{1J}) \exp(x'\alpha_x) + \eta^{-1}\Lambda(w; \beta_{1J}) \exp(x'\beta_x) + \eta^{-1}\Lambda(y - w; \tau_{1J}) \exp(h'\tau_h) + 1\}^{-\eta-2}. \quad (5.16)$$

Turning to treatment effects, we present two options (recall (5.12) and (5.13)).

The first has to do with integrating out δ :

$$E(Y^{+*}|X) = E\{E(Y^{+*}|\delta, X) | X\} = \int_0^\infty \{\eta^{-1}\Lambda(y; \tau_{1J}) \exp(X'\tau_x) + 1\}^{-\eta} dy$$

$$E(Y^{o*}|X) = E\{E(Y^{o*}|\delta, X) | X\} = \int_0^\infty \{\eta^{-1}\Lambda(y; \beta_{1J}) \exp(X'\beta_x) + 1\}^{-\eta} dy.$$

Hence, the mean effect $\Delta_{p\delta}$ with X integrated out and an estimator $\hat{\Delta}_{p\delta}$ for $\Delta_{p\delta}$ are

$$\Delta_{p\delta} \equiv \frac{1}{N} \sum_i \int_0^\infty \{\eta^{-1}\Lambda(y; \tau_{1J}) \exp(X'_i\tau_x) + 1\}^{-\eta} dy - \frac{1}{N} \sum_i \{\eta^{-1}\Lambda(y; \beta_{1J}) \exp(X'_i\beta_x) + 1\}^{-\eta} dy,$$

$$\hat{\Delta}_{p\delta} \equiv \frac{1}{N} \sum_{i=1}^N \frac{1}{n} \sum_{\nu=1}^n [\{\hat{\eta}^{-1}\Lambda(u_\nu; \hat{\tau}_{1J}) \exp(X'_i\hat{\tau}_x) + 1\}^{-\hat{\eta}} - \{\hat{\eta}^{-1}\Lambda(u_\nu; \hat{\beta}_{1J}) \exp(X'_i\hat{\beta}_x) + 1\}^{-\hat{\eta}}]$$

where $u_\nu \sim U[0, \varpi]$, $\nu = 1, \dots, n$, are uniform random numbers with ϖ being the maximum duration in the data.

The other is based on the *proportional hazard-change relative to the untreated hazard*:

$$\Delta_{ph}(t) \equiv \frac{\lambda(t; \tau_{1J})\delta \exp(X'\tau_x) - \lambda(t; \beta_{1J})\delta \exp(X'\beta_x)}{\lambda(t; \beta_{1J})\delta \exp(X'\beta_x)} = \frac{\lambda(t; \tau_{1J})}{\lambda(t; \beta_{1J})} \exp\{X'(\tau_x - \beta_x)\} - 1$$

that is free of δ . The X -integrated sample version to be plotted against t is

$$\hat{\Delta}_{ph}(t) \equiv \frac{\lambda(t; \hat{\tau}_{1J})}{\lambda(t; \hat{\beta}_{1J})} \frac{1}{N} \sum_i \exp\{X'_i(\hat{\tau}_x - \hat{\beta}_x)\} - 1. \quad (5.17)$$

6 Empirical Analysis

6.1 Program Background and Selection into the Program

In this section we apply our estimators to find the effects of an employment subsidy (ES) program administered by the Swedish Public Employment Service (PES). The ES program was introduced on January 1, 1998. The subsidy was targeted at the long-term unemployed, i.e., individuals registered as unemployed at PES for at least 12 months. The subsidy amounted to 50 percent of the total wage and was paid for 6 months at maximum.

In order to receive unemployment insurance (UI) or cash assistance (CA), the unemployed need to register at PES. Thus all long-term unemployed are registered at PES. Whereas the primary objective of PES is to help the long-term unemployed with job search, a secondary objective is to provide training and subsidized employment. Officials at PES may require the unemployed to take a (subsidized) employment or to take part in a training program, and refusal may lead to losing the UI or CA.

The selection-on-observables A_o would hold if the information in X is detailed; in case A_o does not hold, A_u allows the violation of A_o through the single unobserved heterogeneity factor δ . We discuss in the following whether it is reasonable to assume that selection into treatment (more precisely, the waiting duration W) is independent of the potential durations Y^{o*} and Y^{w*} after conditioning on X (and δ).

Using stated-preference experimental data, Eriksson (1997) found that the heterogeneity of the PES case-workers is more important for determining the program participation than the heterogeneity of the individuals. Carling and Richardson (2004) also reported evidence in the same vein using observational data: they compared the effects of eight different programs on the probability of finding a job to show that program placement depended more on the employment service office than on the observed characteristics of the unemployed. There is also survey evidence on the selection process (Lundin, 2000) with the survey directed to the case-workers at PES: only 6 percent of the case-workers maintained that the initiative for ES came from the unemployed.

The above evidence suggests that individual self-selection to ES is not a big prob-

lem, and that PES office characteristics matter much. Ideally, we would like to control for the PES offices in our analysis. This is not possible, however, in our data; instead, we control for local labor market conditions. The local labor market and the PES offices overlap to some extent; there were about 300 PES offices and 100 local labor markets defined on the basis of commuting patterns. Overall, given that our data have detailed information on the labor market history of the individuals and that we can control for the local labor market where the individual is registered, the assumption A_o (or A_u at minimum) looks plausible.

6.2 Data

We use register data from the National Labor Market Board. This unemployment register contains information on all individuals registered at PES in Sweden since August 1991. The database includes detailed information on, e.g., age, educational attainment, sex, registration date, job training activities, and starting dates of participation in various labor market programs.

For each individual registered at PES, we observe the event history including the number of spells and days of unemployment. Anyone who left the register before the introduction of the ES program or whose first spell of unemployment occurred before January 1992 was removed, because the previous labor market history that is an important determinant for unemployment durations and program participation is not available for those individuals. Also, all individuals with the register spells shorter than 365 days were excluded, because the main eligibility criterion for the program is continuous unemployment for at least 365 days.

We focus on individuals of age 25-63 at the time of registration. A spell of unemployment is defined as an uninterrupted period of time when an unemployed person is registered at PES. The spell ends “with employment” if the person finds a job for at least 30 days, and “with censoring” if he/she leaves the register for at least 30 days for any other reason or if the spell goes over the sampling end point (October 1, 2002).

We aggregate the daily data to monthly intervals. The reason for this is that there are measurement errors in the starting day of a job spell. This happened due to

the strategy used by the PES offices to obtain the information on when the job spell began: if the unemployed individual has not been in contact with the PES office for some specific time period, the person is asked over the phone whether employed or not.

Variable	Description	ES (3%)	No ES (97%)
W	Duration to treatment	23.3
Y	Duration	34.38	23.37
Q	=1 if not censored	0.64	0.39
Male	=1 if male	0.61	0.41
Nordic	=1 if Nordic citizen	0.79	0.86
UI	=1 if unemployment insurance	0.84	0.82
Disabled	=1 if disabled	0.06	0.10
Gymnasium	=1 if upper secondary degree	0.43	0.35
University	=1 if university degree	0.12	0.12
Age ₂₀	=1 if age ≤ 30	0.26	0.22
Age ₃₀	=1 if $30 < \text{age} \leq 40$	0.32	0.31
Age ₄₀	=1 if $40 < \text{age} \leq 50$	0.27	0.24
TD ₁	=1 if days in register during previous spell (TD) is 0	0.41	0.38
TD ₂	=1 if $0 < \text{TD} \leq 100$	0.05	0.05
TD ₃	=1 if $100 < \text{TD} \leq 500$	0.22	0.20
TD ₄	=1 if $500 < \text{TD} \leq 1000$	0.18	0.18
TP ₁	=1 if past # programs (TP) is 0	0.41	0.38
TP ₂	=1 if $0 < \text{TP} \leq 5$	0.42	0.39
TP ₃	=1 if $5 < \text{TP} \leq 15$	0.17	0.22

The total number of persons aged 25-63 who were eligible for the ES program during January 1 1998 to October 1 2002 was 631,358, and they are described in Table 1 in two groups: those who started the ES program, and those who did not. Only 3.2% of the eligible spells ended up with ES, and thus the control group is more than

30 times greater than the treatment group, which means that the descriptive statistics for the two groups pooled is almost the same as that of the control group.

A significantly higher fraction (64% v. 39%) of the ES participants ended up in employment. This, however, does not necessarily indicate a good treatment effect, as it may be due to the participants getting registered earlier to spend a longer time looking for a job. The ES participants have higher proportions of males and non-Nordic individuals, and a lower proportion of the disabled. The education level is higher among the ES participants, and the ES participants spent less time registered during the last unemployment period.

6.3 Estimation Results

In specifying the baseline hazards, our main concern is the ‘lock-in effect’ (an initial decline in hazard while the ES lasts) for the treated duration Y^{w*} ; for job trainings, lock-in effects can occur during the training period which hinders job search activity. Since Y^{w*} appears in the likelihood functions only in the form $y_w \equiv y - w$, we specify the baseline hazard $\phi(\cdot)$ for Y^{w*} as

$$\phi(y_w) \equiv 1[y_w \leq 3] + \tau_1 1[3 < y_w \leq 6] + \tau_2 1[6 < y_w \leq 12] + \tau_3 y_w^{\tau_3 - 1} 1[12 < y_w].$$

This specification is primarily due to the time profile of ES lasting 6 months. We set the first threshold at 3, not at 6, to check for the specification, and the last threshold 12 is an admittedly arbitrary choice for the timing of the tail Weibull hazard. Since W and Y^{o*} are not subject to the ES-kind of institutional constraints, we use the Weibull hazard for W and Y^{o*} , which is also to keep the number of parameters small.

The estimated parameters (t-values) are in Table 2; although the region dummies were used, their results are not shown. Due to the huge sample size, most estimates are statistically significant; thus the ‘economic significance’ reflected in the estimate magnitude may be given more importance. Columns (1)-(3) show the hazard parameter estimates for W , Y^{o*} and Y^{w*} . Column (4) for Y^{w*} presents the estimates for the terms interacting with w as well as the estimates for w , w^2 and w^3 . The last row shows $\eta = 45$ for δ , from which we obtain $SD(\delta) = \sqrt{1/45} \simeq 0.15$, giving a 99% confidence interval

$1 \pm 2.58 \times 0.15 \simeq 1 \pm 0.4$ for δ . This means that the unobserved heterogeneity can make the hazard as much as 40% lower/higher than the average for some people.

Table 2: Estimates (excluding region dummies) and t-value				
	(1): α for W	(2): β for Y^{o*}	(3): τ_x for Y^{w*}	(4): τ_{mw} for Y^{w*}
τ_1			2.74 (30.3)	
τ_2			3.42 (35.0)	
$\alpha_0, \beta_0, \tau_3$	1.08 (614)	1.10 (1524)	1.44 (115)	
1	-7.24 (-193)	-4.00 (-446)	-4.26 (-46)	
Male	0.81 (55.6)	-0.09 (-26.6)	0.02 (0.66)	-0.02 (-1.46)
Nordic	-0.61 (-31.2)	-0.21 (-41.2)	-0.06 (-1.27)	0.00 (-0.12)
UI	0.05 (2.24)	-0.54 (-120)	-0.28 (-4.91)	-0.03 (-1.43)
Disabled	-0.85 (-28.9)	-0.06 (-12.4)	0.08 (1.24)	0.01 (0.40)
Gymnasium	0.47 (29.7)	-0.13 (-31.7)	-0.27 (-6.10)	0.05 (3.62)
University	0.34 (14.2)	-0.08 (-13.4)	-0.40 (-5.58)	0.09 (3.47)
AGE ₂₀	0.06 (3.03)	0.06 (12.0)	-0.03 (-0.54)	-0.03 (-1.70)
AGE ₃₀	0.07 (3.51)	-0.07 (-13.4)	-0.13 (-2.33)	-0.02 (-1.31)
AGE ₄₀	-0.37 (-15.3)	0.25 (48.0)	-0.11 (-1.92)	0.02 (1.13)
TD ₁	-0.64 (-7.76)	0.51 (30.0)	-0.67 (-2.17)	-0.01 (-0.15)
TD ₂	-0.49 (-6.30)	0.67 (41.9)	-0.79 (-2.70)	0.06 (0.67)
TD ₃	-0.43 (-5.46)	0.86 (52.9)	-1.00 (-3.36)	0.14 (1.47)
TD ₄	-0.51 (-6.34)	1.31 (80.0)	-0.97 (-3.26)	0.18 (1.87)
TP ₁	0.87 (11.2)	-0.44 (-27.9)	0.68 (2.32)	0.02 (0.17)
TP ₂	1.04 (13.1)	-0.03 (-2.05)	0.83 (2.80)	0.09 (0.92)
TP ₃	1.36 (9.93)	0.45 (17.1)	1.59 (4.19)	1.15 (6.07)
w				-0.21 (4.21)
$w^2/100$				0.013 (1.14)
$w^3/10000$				0.030 (3.41)
η for δ			44.9 (24.0)	

For Column (1), α_0 shows a small increasing hazard to ES. Male, non-Nordic origin, no disability, high education, and ‘age higher than 50 relative to age 40-50’ raise the

hazard to ES. Relative to the base category of more than 1000 days in register during the previous spell, fewer days in register decrease the hazard to ES, which is shown by TD_j , $j = 1, 2, 3, 4$. Compared with the base category of the previous number of programs being greater than 15, fewer numbers increase the hazard to ES.

As for Column (2), β_0 shows a small increasing hazard to employment without getting treated. Nordic origin, UI and high education decrease the hazard. Relative to the base category of age 50+, individuals aged 40-50 have a higher hazard. Compared with the base category of more than 1000 days in register during the previous spell, fewer days in register increase the hazard.

Regarding Column (3), judging from τ_1 and τ_2 , the hazard over $(3, 6]$ increases the hazard over $(0, 3]$ by almost 3 times whereas the hazard increases by only 25% from 2.74 to 3.42 when ES runs out at $t = 6$. As was the case for Y^{o*} , UI and high education decreases the hazard to employment with the treatment. But differently from Y^{o*} , compared with the base category of more than 1000 days in register during the previous spell, fewer days in register decrease the hazard. Relative to the base category of the previous number of programs being greater than 15, fewer numbers increase the hazard.

Column (4) shows that only education and TP_3 dummies interacting with w are significant with positive estimates. As for the cubic function of w , the linear and cubic terms are significant. Although w has a negative slope while the other two terms have positive slopes, w dominates the other terms so that the entire w function is negative; i.e., the waiting time to ES decreases the hazard (i.e., increases the treated duration) to result in $Y^{w*} > Y^{+*}$. In fact, the cubic function is almost the same as $-0.16w$ for $w \in [0, 100]$.

Recalling (5.17) and its preceding equations, we obtain the average effect estimator $\hat{\Delta}_{p\delta}$ using $\varpi = 57$ that is the sample maximum duration for the treatment group. We also obtain $\hat{\Delta}_{ph}(t)$ for the proportional hazard difference. For both estimators, the standard deviations are calculated by the delta method:

$$SD\{\hat{\Delta}_z(\theta_z)\} = \{c(\hat{\theta}_z)' \cdot \hat{V}_z \cdot c(\hat{\theta}_z)\}^{1/2} \quad \text{where } c(\hat{\theta}_z) \equiv \frac{\partial \hat{\Delta}_z(\theta_z)}{\partial \theta_z} \Big|_{\theta_z = \hat{\theta}_z} \text{ for } z = ph, p\delta$$

where $\theta_{ph} \equiv (\beta_0, \beta'_x, \tau_1, \tau_2, \tau_3, \tau'_x)'$, $\theta_{p\delta} \equiv (\theta'_{ph}, \eta)'$ and \hat{V}_z is the asymptotic variance matrix estimator for $\hat{\theta}_z$.

The result for $\hat{\Delta}_{p\delta}$ with $\varpi = 57$ is

$$\hat{\Delta}_{p\delta} = -11.1 \quad \text{with } SD\{\hat{\Delta}_{p\delta}(\theta_{p\delta})\} = 1.37 \text{ and t-value } -8.09 :$$

ES reduces unemployment duration by 11.1 months on average. This estimate is close to the estimates in Fredriksson and Johansson (2008). Incidentally, -11.1 is also close to $(\bar{Y}^1 - \bar{W}) - \bar{Y}^0 = (34.4 - 23.3) - 23.4 = -12.3$ from Table 1 as if there were neither censoring nor covariate.

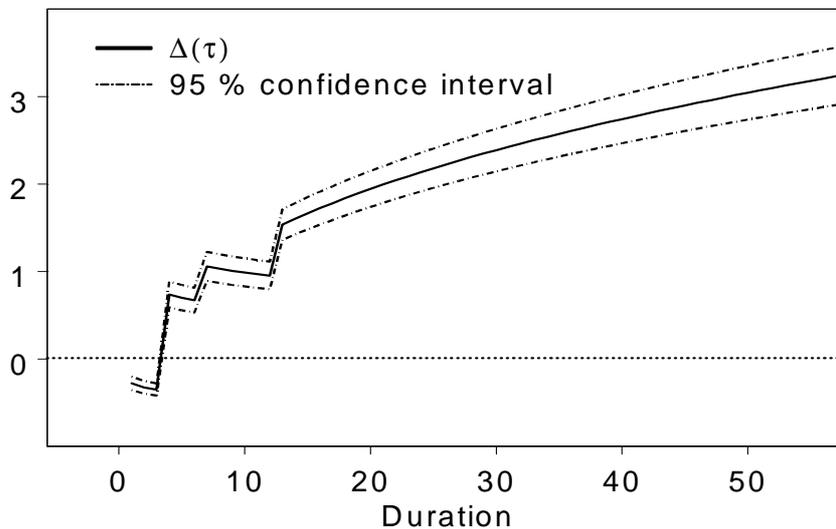


Figure 1: Proportional Hazard Difference

The results for $\hat{\Delta}_{ph}(t)$ is in Figure 1 along with the 95% point-wise asymptotic confidence interval. Figure 1 shows a lock-in effect for the first three months in the ES program as the graph shows negative values, and thereafter, we find a higher hazard of leaving unemployment due to ES. Had we used piecewise constant hazards for all durations, the graph would be step-shaped; since a piecewise constant hazard was used only for Y^{w*} while a Weibull hazard was used for Y^{o*} , the graph shows slightly declining patterns on each “step”.

7 Conclusions

In observational data, a treatment/program is chosen at some time point by the subject involved, and when it is chosen at that particular time point, it must be chosen for a good reason. Despite this, the treatment timing choice is usually ignored, and the data set at hand is divided into treatment and control groups based on a time interval; with a longer interval, however, many control subjects become treated, leading to possibly much different findings.

In this paper, we set up an “ideal” causal framework for finding treatment effects on durations, taking into account the treatment-timing choice—ideal in the sense that the causal framework takes the treatment-timing aspect out of the picture. We envisioned three durations (waiting time to treatment, untreated duration from the baseline, and treated duration from the treatment timing), and the main causal effect was defined as the difference between the counter-factual treated duration at the baseline and the untreated duration at the baseline.

Initially we assumed that the three durations are independent of one another given the observed covariates, which is a selection-on-observable approach. We then introduced a multiplicative unobserved heterogeneity term (“factor”) so that the three durations are allowed to be related through the common factor, which is a selection-on-unobservable approach. Although restrictive, by assuming that the factor follows gamma distribution, we were able to obtain a close-form likelihood function despite the three durations entangled in complex ways. To avoid parametric assumptions, we used piecewise-constant hazards as well as Weibull hazards.

With the unobserved factor and piecewise-constant hazards in, a two-stage version of the MLE estimating the waiting and untreated duration parameters first and then the treated duration parameters later converged well; this MLE is our simple answer to the complicated treatment timing problem. We applied our methods to Swedish employment subsidy data to find effects similar to those found by purely nonparametric methods in the literature.

APPENDIX

Derivation of Four Likelihood Components (4.1)-(4.4)

(4.1) to (4.4) follow from, respectively,

$$\begin{aligned} P(D = 0, Q = 0|C = y, X)f_{C|X}(y) &= P(C < W, C < Y^{o*}|C = y, X)f_{C|X}(y) \\ &= P(y < W|C = y, X) \cdot P(y < Y^{o*}|C = y, X) \cdot f_{C|X}(y) = P(y < W|X)P(y < Y^{o*}|X)f_{C|X}(y); \end{aligned}$$

$$\begin{aligned} P(D = 0, Q = 1|Y^{o*} = y, X)f_{Y^{o*}|X}(y) &= P(Y^{o*} < W, Y^{o*} < C|Y^{o*} = y, X)f_{Y^{o*}|X}(y) \\ &= P(y < W|Y^{o*} = y, X) \cdot P(y < C|Y^{o*} = y, X) \cdot f_{Y^{o*}|X}(y) = P(y < W|X)P(y < C|X)f_{Y^{o*}|X}(y); \end{aligned}$$

$$\begin{aligned} P(D = 1, Q = 0|W = w, C = y, X)1[w < y] \cdot f_{W|X}(w)f_{C|X}(y) \\ &= P(W < Y^{o*}, C - W < Y^{w*}|W = w, C = y, X)1[w < y] \cdot f_{W|X}(w)f_{C|X}(y) \\ &= P(w < Y^{o*}|W = w, C = y, X) \cdot P(y - w < Y^{w*}|W = w, C = y, X)1[w < y]f_{W|X}(w)f_{C|X}(y) \\ &= P(w < Y^{o*}|X) \cdot P(y - w < Y^{w*}|X)1[w < y] \cdot f_{W|X}(w)f_{C|X}(y); \end{aligned}$$

$$\begin{aligned} P(D = 1, Q = 1|W = w, Y^{w*} = y - w, X) \cdot 1[w < y]f_{W|X}(w)f_{Y^{w*}|X}(y - w) \\ &= P(W < Y^{o*}, W + Y^{w*} < C|W = w, Y^{w*} = y - w, X) \cdot 1[w < y]f_{W|X}(w)f_{Y^{w*}|X}(y - w) \\ &= P(w < Y^{o*}|W = w, Y^{w*} = y - w, X) \cdot P(y < C|W = w, Y^{w*} = y - w, X) \\ &\quad \cdot 1[w < y]f_{W|X}(w)f_{Y^{w*}|X}(y - w) = P(w < Y^{o*}|X)P(y < C|X)1[w < y]f_{W|X}(w)f_{Y^{w*}|X}(y - w). \end{aligned}$$

Three integrals (5.1), (5.2) and (5.3)

$$\begin{aligned} E\{\exp(-pZ)\} &= \int \exp(-pz) \frac{\eta^\eta}{\Gamma(\eta)} z^{\eta-1} e^{-\eta z} dz = \frac{\eta^\eta}{\Gamma(\eta)} \int \exp\{-(p + \eta)z\} z^{\eta-1} dz \\ &= \frac{\eta^\eta}{\Gamma(\eta)} \int e^{-c} \left(\frac{c}{p + \eta}\right)^{\eta-1} \frac{1}{p + \eta} dc \quad (\text{setting } c = (p + \eta)z \implies dz = \frac{1}{p + \eta} dc) \\ &= \frac{1}{\Gamma(\eta)} \left(\frac{\eta}{p + \eta}\right)^\eta \int e^{-c} c^{\eta-1} dc = \frac{1}{\Gamma(\eta)} \left(\frac{\eta}{p + \eta}\right)^\eta \Gamma(\eta) = \left(\frac{1}{\eta^{-1}p + 1}\right)^\eta = (\eta^{-1}p + 1)^{-\eta}; \end{aligned}$$

$$\begin{aligned}
E\{Z \exp(-pZ)\} &= \int z \exp(-pz) \frac{\eta^\eta}{\Gamma(\eta)} z^{\eta-1} e^{-\eta z} dz = \frac{\eta^\eta}{\Gamma(\eta)} \int \exp\{-(p+\eta)z\} z^\eta dz \\
&= \frac{\eta^\eta}{\Gamma(\eta)} \int e^{-c} \left(\frac{c}{p+\eta}\right)^\eta \frac{1}{p+\eta} dc = \frac{1}{\Gamma(\eta)} \frac{\eta^\eta}{(p+\eta)^{\eta+1}} \int e^{-c} c^\eta dc \\
&= \frac{1}{\Gamma(\eta)} \frac{\eta^\eta \cdot \Gamma(\eta+1)}{(p+\eta)^{\eta+1}} = \frac{1}{\Gamma(\eta)} \frac{\eta^\eta \cdot \eta \Gamma(\eta)}{(p+\eta)^{\eta+1}} = \left(\frac{\eta}{p+\eta}\right)^{\eta+1} = (\eta^{-1}p+1)^{-\eta-1};
\end{aligned}$$

$$\begin{aligned}
E\{Z^2 \exp(-pZ)\} &= \int z^2 \exp(-pz) \frac{\eta^\eta}{\Gamma(\eta)} z^{\eta-1} e^{-\eta z} dz = \frac{\eta^\eta}{\Gamma(\eta)} \int \exp\{-(p+\eta)z\} z^{\eta+1} dz \\
&= \frac{\eta^\eta}{\Gamma(\eta)} \int e^{-c} \left(\frac{c}{p+\eta}\right)^{\eta+1} \frac{1}{p+\eta} dc = \frac{1}{\Gamma(\eta)} \frac{\eta^\eta}{(p+\eta)^{\eta+2}} \int e^{-c} c^{\eta+1} dc = \frac{1}{\Gamma(\eta)} \frac{\eta^\eta \cdot \Gamma(\eta+2)}{(p+\eta)^{\eta+2}} \\
&= \frac{1}{\Gamma(\eta)} \frac{\eta^\eta \cdot (\eta+1)\eta \Gamma(\eta)}{(p+\eta)^{\eta+2}} = \frac{\eta+1}{\eta} \left(\frac{\eta}{p+\eta}\right)^{\eta+2} = (1+\eta^{-1}) \cdot (\eta^{-1}p+1)^{-\eta-2}.
\end{aligned}$$

Derivation for the δ -free likelihoods (5.10) and (5.11)

For $(D=1, Q=0)$, recalling (4.5), the δ -conditional likelihood for $(W=w, C=y)$ is

$$\begin{aligned}
&S_{Y^{0*}|X=x,\delta}(w) S_{Y^{w*}|X=x,\delta}(y-w) \cdot 1[w < y] f_{W|X=x,\delta}(w) \\
&= \exp\{-w^{\beta_0} \delta \exp(x' \beta_x) - (y-w)^{\tau_0} \delta \exp(h' \tau_h)\} \cdot 1[w < y] \\
&\cdot \alpha_0 w^{(\alpha_0-1)} \delta \exp(x' \alpha_x) \exp\{-w^{\alpha_0} \delta \exp(x' \alpha_x)\} \\
&= 1[w < y] \alpha_0 w^{(\alpha_0-1)} \exp(x' \alpha_x) \delta \exp[-\{w^{\alpha_0} \exp(x' \alpha_x) + w^{\beta_0} \exp(x' \beta_x) + (y-w)^{\tau_0} \exp(h' \tau_h)\} \delta].
\end{aligned}$$

Use (5.2) with $p = w^{\alpha_0} \exp(x' \alpha_x) + w^{\beta_0} \exp(x' \beta_x) + (y-w)^{\tau_0} \exp(h' \tau_h)$ to obtain (5.10).

For $(D=1, Q=1)$, recalling (4.6), the δ -conditional likelihood for $(W=w, W+Y^{W*}=y)$ is

$$\begin{aligned}
&S_{Y^{0*}|X=x,\delta}(w) \cdot 1[w < y] f_{W|X=x,\delta}(w) f_{Y^{w*}|X=x,\delta}(y-w) \\
&= \exp\{-w^{\beta_0} \delta \exp(x' \beta_x)\} \cdot 1[w < y] \cdot \alpha_0 w^{(\alpha_0-1)} \delta \exp(x' \alpha_x) \exp\{-w^{\alpha_0} \delta \exp(x' \alpha_x)\} \\
&\cdot \tau_0 (y-w)^{\tau_0-1} \delta \exp(h' \tau_h) \exp\{-(y-w)^{\tau_0} \delta \exp(h' \tau_h)\} \\
&= \alpha_0 w^{(\alpha_0-1)} \exp(x' \alpha_x) \cdot \tau_0 (y-w)^{\tau_0-1} \exp(h' \tau_h) \cdot 1[w < y] \\
&\cdot \delta^2 \exp[-\{w^{\alpha_0} \exp(x' \alpha_x) + w^{\beta_0} \exp(x' \beta_x) + (y-w)^{\tau_0} \exp(h' \tau_h)\} \delta].
\end{aligned}$$

Using (5.3), integrate δ out to obtain (5.11).

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, 2007, The unobserved heterogeneity distribution in duration analysis, *Biometrika* 94, 87-99.

Carling, K. and K. Richardson, 2004, The relative efficiency of labor market programs: Swedish experience from the 1990s, *Labour Economics* 11, 335-354.

De Luna, X. and P. Johansson, 2010, Nonparametric inference for the effect of a treatment on survival times with application in the health and social sciences, *Journal of Statistical Planning and Inference* 140, 2122-2137.

Eberwein, C., J.C. Ham and R.J. Ralonde, 1997, The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: evidence from experimental data, *Review of Economic Studies* 64, 655-682.

Eriksson, M., 1997, Placement of unemployed into labour market programs: a quasi experimental study, *Umeå Economic Studies* 439, Umeå university.

Fredriksson, P. and P. Johansson, 2008, Dynamic treatment assignment: the consequences for evaluations using observational data, *Journal of Business and Economic and Statistics* 26, 435-445.

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.

Han, A. and J.A. Hausman, 1990, Flexible parametric estimation of duration and competing risk models, *Journal of Applied Econometrics* 5, 1-28.

Heckman, J.J. and B. Singer, 1984, A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica* 52, 271-320.

Imbens, G.W. and J.M. Wooldridge, 2009, Recent developments in the econometrics of program evaluation, *Journal of Economic Literature* 47, 5-86.

Lancaster, T., 1979, Econometric methods for the duration of unemployment, *Econometrica* 47, 939-56.

Lee, M.J., 2005, *Micro-econometrics for policy, program, and treatment effects*, Oxford University Press.

Lillard, L.A., 1993, Simultaneous equations for hazards, *Journal of Econometrics* 56, 189-217.

Lundin, M., 2000, *Anställningsstödens implementering vid arbetsförmedlingarna*, Stencilserie 2000:4, IFAU, Uppsala.

Pearl, J., 2009, *Causality*, 2nd ed., Cambridge University Press.

Rosenbaum, P.R., 2002, *Observational studies*, 2nd ed., Springer.