



van den Berg, G. J., Bonev, P., & Mammen, E. (2020). Nonparametric Instrumental Variable Methods for Dynamic Treatment Evaluation. *Review of Economics and Statistics*, 102(2), 355-367.  
[https://doi.org/10.1162/rest\\_a\\_00843](https://doi.org/10.1162/rest_a_00843)

Peer reviewed version

License (if available):  
Other

Link to published version (if available):  
[10.1162/rest\\_a\\_00843](https://doi.org/10.1162/rest_a_00843)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) will be available online via The MIT Press. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# NONPARAMETRIC INSTRUMENTAL VARIABLE METHODS FOR DYNAMIC TREATMENT EVALUATION

GERARD J. VAN DEN BERG\*, PETYO BONEV†, ENNO MAMMEN‡

ABSTRACT. We develop an instrumental variable approach for identification of dynamic treatment effects on survival outcomes in the presence of dynamic selection, noncompliance and right-censoring. The approach is nonparametric and does not require independence of observed and unobserved characteristics or separability assumptions. We propose estimation procedures and derive asymptotic properties. We apply our approach to evaluate a policy reform in which the pathway of unemployment benefits as a function of the unemployment duration is modified. Those who were unemployed at the reform date could choose between the old and the new regime. We find that the new regime has a positive average causal effect on the job finding rate.

---

Keywords: dynamic treatment effects, noncompliance, instrumental variable, unemployment duration, active labor market policies.

JEL codes: C14, C41, J64, J65.

\* University of Bristol, IFAU-Uppsala, IZA, ZEW, J-PAL..

†University of St. Gallen. Email: petyo.bonev@unisg.fr.

‡Heidelberg University and National Research University Higher School of Economics, Russian Federation. Email: mammen@math.uni-heidelberg.de.

We thank the Editor, three anonymous Referees as well as Sylvie Blasco, Christoph Breunig, Bettina Drepper, Markus Frölich, Bo Honoré, Andreas Landmann, Aureo de Paula, Gautam Tripathi and participants at the ESEM, an IZA conference on labor market policy evaluation at Harvard, conferences on survival analysis and on the evaluation of political reforms at Mannheim, a workshop at ZEW, and the joint econometrics and statistics workshop at the LSE, for their useful comments. We thank INSEE-CREST and DARES at the French Ministry of Labor, especially Bruno Crépon, Thomas le Barbanchon, Francis Kramarz, and Philippe Scherrer, for their extraordinary help with the data access and their hospitality and for having shared their institutional and econometric expertise.

## 1. INTRODUCTION

In the evaluation of labor market policies, such as job search assistance and classroom training, it is usually interesting to compare the impact of the policy on the long-term unemployed to the impact on those who only lost their jobs very recently. Differences between effects at low durations and high durations may shed light on the extent to which individual behavior changes over time and this may be relevant for policy design (see e.g. [Van den Berg \(2001\)](#)). Empirical and theoretical studies therefore tend to focus on the evaluation of dynamic treatment effects conditional on survival at a range of elapsed durations.

However, the identification of such dynamic treatment effects is typically hampered by several hurdles. First, if individuals can choose a treatment arm different from the one assigned to them (noncompliance), then standard conditional independence assumptions will be violated. Second, suppose the treatment is randomized at the inflow into unemployment. In the presence of unobserved determinants of the outcome, their distributions among survivors at some later point in time may differ across different treatment arms (dynamic selection), see [Meyer \(1996\)](#), [Ham and LaLonde \(1996\)](#), [Eberwein et al. \(1997\)](#) and [Abbring and van den Berg \(2005\)](#). This raises the question of how to choose the treatment and control groups in a dynamic setup. Finally, when the outcome of interest is a duration variable, identification might be hampered by right-censoring.

In this paper, we develop an instrumental variable (IV) approach for identification of dynamic treatment effects in the presence of dynamic selection, noncompliance and right-censoring. Our method is fully nonparametric and we do not impose independence of observed and unobserved characteristics or separability in their effects on the outcome. We propose estimation procedures and derive their asymptotic properties.

At the core of our method is a dynamic potential outcomes framework. A treatment is assigned at some random elapsed duration of unemployment. The interest is in the effect of this treatment on post-treatment outcomes such as post-treatment unemployment duration. A major question in this setup is how to define a meaningful treatment effect. While the standard static literature defines those who are not observed to enroll into the treatment as nontreated, in a dynamic setting this approach leads to conditioning on future successful outcomes, [Sianesi \(2004\)](#).<sup>1</sup> To avoid this problem, we follow much of the literature on dynamic treatment effects and focus on treatment effects on the subgroup of those individuals who remain unemployed at least until the treatment, see e.g. [Sianesi \(2004\)](#), [Vikström \(2014\)](#) and [Van den Berg et al. \(2014\)](#).

Our main contribution is to extend the dynamic treatment evaluation framework to allow for endogenous noncompliance. Noncompliance has been largely ignored by the literature on dynamic treatment effects, despite the attention it has received in the static literature. We develop a dynamic concept of noncompliance which allows the individual to change preferences in an arbitrary way over time. The intuition behind is that whether an unemployed accepts to participate or not in a labor market program depends on her subjective probability to find a job and this probability might change with the elapsed duration of unemployment.

Our method relies on two major assumptions. The first one is exogenous variation in the timing of treatment assignment. As a motivation, consider a case worker who is responsible for a large pool of unemployed. Then, conditional on characteristics of the unemployed, the case worker might assign the order in which the unemployed are advised idiosyncratically, see e.g. [Sianesi \(2004\)](#). Our strategy is also motivated by the so-called phased-in experiments, in which randomly selected late recipients provide a

---

<sup>1</sup> In particular, individuals who find a job before the treatment is assigned to them are considered as nontreated.

control group for randomly selected early recipients, see [Duflo et al. \(2007\)](#). Finally, when a policy is administered at a single point in time, then the presence of cohorts who enter unemployment at different points in time might also give rise to quasi-experimental variation in the time to treatment.

The second one is a dynamic consistency assumption commonly referred to as “no anticipation” in the literature, see [Abbring and van den Berg \(2003\)](#). This assumption requires that if two labor market treatments coincide up to a duration  $t$ , then the hazards of unemployment duration under the two treatments should be the same for each pre-treatment duration. With forward-looking individuals, this may require that individuals do not anticipate the (time of) treatment or do not act upon their knowledge. The no anticipation assumption has been used throughout the literature on dynamic treatment effects, [Crépon et al. \(2009\)](#), [Crepon et al. \(2010\)](#), [Vikström \(2014\)](#). Moreover, it is an implicit assumption in some standard static approaches, such as the Difference-in-Difference (DiD) and the Synthetic Control approaches. It is also an implicit assumption in the phased-in experiments, where ignoring the time dimension effectively subsumes the no anticipation assumption into the randomization assumption. Our paper puts phased-in experiments into a dynamic framework and makes the link to no anticipation explicit.

Our identification strategy consists of two steps: a dynamic and a static one. In the dynamic step, initial randomization of the assignment and no anticipation ensure that dynamic selection follows the same pattern for early and late treatment recipients at each pre-treatment duration.<sup>2</sup> This allows for a comparison of treated and not-yet-treated individuals at the same elapsed duration of unemployment. In a second step, the assignment

---

<sup>2</sup> [Van den Berg et al. \(2014\)](#) derive this result in a context with full compliance. We generalize their result to a setup with endogenous noncompliance.

to treatment is used as an instrument for the endogenous enrollment into treatment. Information on observed compliance from the early recipients is used to identify the outcome distribution of late, not-yet-treated, recipients.

With these two steps, our paper provides a link between the dynamic treatment evaluation literature, [Vikström \(2014\)](#), [Sianesi \(2004\)](#), [Michael et al. \(2011\)](#), [Van den Berg et al. \(2014\)](#) and the standard static LATE literature, [Imbens and Angrist \(1994\)](#), [Imbens and Rubin \(1997\)](#). Identification is local in the sense that at each point in time the effect is identified only for those who would comply at this particular point in time. The static notion of location is thus extended with a time component. The corresponding estimators can be viewed as dynamic Wald estimators. Moreover, in a setup where time to treatment in the control group approaches the time to treatment in the treatment, our method can be extended to a dynamic fuzzy regression discontinuity (RD) approach.

Our paper contributes to the literature on IV in survival analysis, see e.g. [Eberwein et al. \(1997\)](#), [Robins and Tsiatis \(1991\)](#), [Chesher \(2002\)](#), [Bijwaard and Ridder \(2005\)](#), [Bijwaard \(2008\)](#) and [Tchetgen et al. \(2014\)](#). Much of this literature is surveyed by [Abbring and van den Berg \(2005\)](#). Typically, these studies adopt a semiparametric or a parametric model structure. Our model, on the contrary, is fully nonparametric. Dynamic discrete choice models also deal with identification of dynamic treatment effects, see e.g. [Heckman and Navarro \(2007\)](#). Contrary to our approach, those papers rely on period-specific exclusion restrictions as well as on restrictive separability and identification-at-infinity assumptions. Our approach is also related to the literature on duration models with a Mixed Proportionate Hazard (MPH) structure and time-varying covariates, in particular to the important paper of [Hausman and Woutersen \(2014\)](#). A thorough overview of this literature is provided in [Hausman and Woutersen \(2008\)](#). Similar to our approach, identification in [Hausman and Woutersen \(2014\)](#) relies on variation in the time to treatment. The major difference is how dynamic selection is handled. While we handle dynamic selection

by assuming no anticipation of the treatment, dynamic selection is modeled explicitly in [Hausman and Woutersen \(2014\)](#) through the semi-parametric assumption on the hazard, and in particular through a separability assumption on the unobserved heterogeneity. In addition, their rank estimator utilizes the monotonicity of the hazard function in the observed covariates  $X$  implied by a parametric assumption. Another paper that utilizes variation in the time to treatment as a source of identification is [Abbring and van den Berg \(2003\)](#). Contrary to our nonparametric model, both the treatment process and the effect of the treatment on the outcome duration are modeled within the semiparametric MPH framework. Finally, our dynamic RD extension is related to the static RD approach in [Hahn et al. \(2001\)](#).

An additional contribution of our paper is to develop a theoretical framework for the analysis of noncompliance in a dynamic setting. Specifically, we propose how to test for endogenous noncompliance and how to measure the bias that would arise if endogeneity is ignored. Measuring selection bias can provide valuable insights on the reasons for the non-take up of a policy reform and thus help improve policy design.

We use our approach to evaluate the French 2001 labor market policy reform PARE. This reform introduced a more generous unemployment benefits system together with more stringent monitoring and training measures. Individuals who were unemployed at the moment of the reform could choose whether to stay in the old regime for the remaining duration of their spell or to enter the new regime immediately. Our results suggest that this policy increased the exit rate out of unemployment. Our findings are supported by an extensive empirical examination of the plausibility of the assumptions.

## 2. ECONOMETRIC FRAMEWORK

**2.1. Treatment effects.** For illustrative purposes, we build our exposition on a labor market example. Suppose we observe a sample of  $n$  individuals who are searching for

a job. As part of Active Labor Market Policies (ALMPs), the unemployed are offered a job search training. The elapsed duration at which the individuals are offered the training might vary across individuals and is denoted by  $Z_i$  for individual  $i = 1 \dots n$ . The unemployed may accept or refuse to participate and the training is offered only once. Allowing for noncompliance mimics unemployment insurance (UI) systems in which ALMPs are not enforced or sanctions for violations are either very mild or come with a low probability. Examples for such UI systems are the Swedish, the French and the Australian ones.<sup>3</sup> This setup also applies to experimental studies, in which the subjects are invited to participate (encouragement design, see e.g. [Duflo et al. \(2007\)](#)) or can simply refuse to participate.

Denote by  $S_i$  the actual pre-treatment duration, that is, the time individual  $i$  spends in unemployment until she receives the training. We focus on the case in which if the individual complies, the treatment must be taken immediately, and thus  $S_i$  coincides with  $Z_i$ ,  $S_i = Z_i$ . If the individual refuses the treatment, she is never treated, which we normalize to  $S_i = \infty$ .

We are interested in the effect of the job search training on the unemployment duration. When defining a treatment effect of interest, there are two important aspects endemic to the dynamic setting. First, the timing of the treatment  $S_i$  might matter: job search training might have different effects on the (total or post-treatment) unemployment duration at different elapsed unemployment durations, see [Abbring and van den Berg \(2003\)](#). We therefore allow the counterfactual outcome to depend on the pre-treatment duration  $S_i$ . In particular, for each  $s \in \mathbb{R}^+ \cup \{\infty\}$ , denote by  $T_i(s)$  the potential duration of unemployment if the treatment was received at an elapsed duration  $s$ . With this notation we implicitly

---

<sup>3</sup> See [Sianesi \(2004\)](#), [Crepon et al. \(2010\)](#) and [Carney and Ramia \(2011\)](#).



impose an exclusion restriction on  $Z_i$ ,

$$(2.1) \quad T_i(s, z) = T_i(s) \quad \text{for each } s, z \in \mathbb{R}^+ \cup \{\infty\}.$$

(2.1) prevents the assignment to treatment from directly impacting the outcome:  $Z_i$  can influence the duration only through the actual pre-treatment duration  $S_i$ . This assumption is plausible in our setting since we restrict  $Z_i$  and  $S_i$  to realize simultaneously from viewpoint of the unemployed. This is in contrast to the case in which the assignment to treatment realizes prior to the treatment.<sup>4</sup>

The second aspect that differs from the static framework is that a simple comparison of treated and nontreated within the treatment definition window possibly leads to a bias.<sup>5</sup> In particular, if an individual finds a job prior to the treatment,  $S_i$  will be censored by  $T_i$  and therefore unobserved. Considering these individuals as nontreated effectively conditions on their future successful outcomes, [Sianesi \(2004\)](#). We therefore follow the approach chosen by much of the literature on dynamic treatment effects and condition on survival in unemployment up to treatment, e.g. [Vikström \(2014\)](#) and [Sianesi \(2004\)](#).<sup>6</sup>

The main object of interest is the post-treatment duration of a particular group of compliers. We model compliance in this dynamic framework in the following way. For each  $z \in \mathbb{R}^+ \cup \{\infty\}$ , denote by  $S_i(z)$  the potential pre-treatment duration of individual  $i$  if she was assigned to receive the treatment at  $z$ . Following the exposition above,  $S_i(z)$  can be either equal to  $z$  or to  $\infty$ .

---

<sup>4</sup>The standard example of non-simultaneous realizations is when the individual is warned about a future sanction due to noncompliance with UI rules, see [Crepon et al. \(2010\)](#). In this case,  $Z_i$  might have a direct effect on the outcome, which is often referred to as a threat effect.

<sup>5</sup>Treatment definition window is the period of time used in a static setup to define the treatment status.

<sup>6</sup>Following the standard terminology in survival analysis, we refer to the remaining in the state of interest (e.g. unemployment) to as survival, and to the corresponding individuals as survivors.

With these preliminaries, we define the treatment effect of interest as

$$(2.2) \quad TE(t, t', a) = \mathbb{E}[P\{T(t) \in [t, t+a] \mid T(t) \geq t, X, V, S(t) = t\} \\ - P\{T(t') \in [t, t+a] \mid T(t') \geq t, X, V, S(t) = t\} \mid T(t) \geq t, X, S(t) = t],$$

where  $t$  and  $t'$  be two fixed elapsed durations with  $t < t'$ ,  $a$  is a positive number in  $(0, t' - t]$ ,  $X$  is an observed random vector of individual characteristics such as age, qualification and experience, and  $V$  is an unobserved random variable that captures ex ante heterogeneity in terms of unobserved characteristics. We may refer to  $V$  as unobserved confounders. In labor market register data, such as our data, these may be non-cognitive abilities such as the degree of intrinsic motivation. Both  $X$  and  $V$  are time-constant and realized prior the spell of unemployment.<sup>7</sup>

The interpretation of (2.2) is as follows. Since  $t + a < t'$ , all individuals assigned to be treated at  $t'$  are not yet treated in  $[t, t + a)$ . Thus,  $P\{T(t') \in [t, t + a] \mid T(t') \geq t, X, V, S(t) = t\}$  can be interpreted as the individual potential outcome of a nontreated individual who would remain unemployed under the treatment  $t'$  at least until  $t$ .  $P\{T(t) \in [t, t + a] \mid T(t) \geq t, X, V, S(t) = t\}$  is the corresponding outcome of a treated at  $t$  individual. Thus,

$$(2.3) \quad P\{T(t) \in [t, t + a] \mid T(t) \geq t, X, V, S(t) = t\} - P\{T(t') \in [t, t + a] \mid T(t') \geq t, X, V, S(t) = t\}$$

is the additive individual treatment effect on the probability to leave unemployment in  $[t, t + a)$  for an individual who is still unemployed at  $t$ . Expression (2.2) is an average of (2.3). The average is built with respect to the distribution of unobserved heterogeneity

---

<sup>7</sup> We suppress the index  $i$  for notational simplicity.

$V$  among the treated survivors,  $F_{V|T(t) \geq t, X, S(t)=t}$ .<sup>8</sup> Proposition 3.1 shows, that under certain assumption this is also the distribution of  $V$  among all survivors at  $t$ .

Conditioning on  $S(t) = t$  restricts the evaluation of the effect on the  $t$ -compliers, i.e. the individuals who would take the treatment at an elapsed duration  $t$  if it was assigned to them. Offering the treatment only once means that we only allow for one-sided noncompliance. With one-sided noncompliance, the set of treated compliers coincides with the full set of treated (at elapsed duration  $t$ ). Therefore, (2.2) can be interpreted as a *treatment effect on the treated*.

By conditioning on survival up to treatment and focusing on the post-treatment outcome, (2.2) closely resembles the definition of a treatment effect in the dynamic impact evaluation literature. The novel aspect which our paper introduces is to explicitly allow for noncompliance. The stochastic process  $\{S(t)\}_{t \geq 0}$  is a generalization of the static compliance model in the LATE literature, [Imbens and Angrist \(1994\)](#). (2.2) thus provides a natural link between the dynamic treatment evaluation literature and the static LATE.

**Remarks. 1.** It is clear from definition (2.2) that the duration outcome  $T(s)$  can be replaced by an arbitrary post-treatment outcome  $Y(s)$ . In fact, our identification strategy presented in the next section does not rely on the outcome being a duration variable.<sup>9</sup> **2.** For any two  $t, t'$ , there is a variety of possible treatment effects, one for each  $a \in (0, t' - t)$ . In addition, an evaluation of the total effect of a policy might involve averages over all  $t$  and  $t'$ . **3.** An important special case of (2.2) is the limit case  $a \rightarrow 0$ . It amounts to a treatment effect on the hazard function. We devote special attention to this case in the next section and in the appendix. **4.** The precise interpretation of (2.2) is less straightforward when

---

<sup>8</sup> Note that (2.2) can be also written as

$$\mathbb{E}_{V|T(t) \geq t, X, S(t)=t} [P\{T(t) \in [t, t+a) \mid T(t) \geq t, X, V, S(t) = t\} - P\{T(t') \in [t, t+a) \mid T(t') \geq t, X, V, S(t) = t\}]$$

<sup>9</sup> Our estimator, however, utilizes the relation between the outcome and the conditioning set.

the treatment is not instantaneous. As an example, long term training measures induce a lock-in effect. One approach followed in the literature is to treat the lock-in effect as constituent of the total treatment effect, see [Sianesi \(2004\)](#). A second problem, however, is that the length of a non-instantaneous treatment might also matter. We therefore concentrate on instantaneous treatments, such as short-term training, counseling and other activating measures. In the context of the labor market example, such focus is not an important restriction, as there is a general trend in labor market policies towards short-term activation and reemployment ALMPs, see e.g. [Biewen et al. \(2014\)](#). 5. Expression (2.2) implies that we treat  $T(s)$  as a random variable even when we condition on all observed and unobserved characteristics  $X, V$ . By doing that, we follow the general approach in mixture duration models pioneered by [Lancaster \(1979\)](#). The underlying assumption is that the randomness in  $T(s)$  comes from some intrinsic uncertainty in the transition, not observed and controlled by the individual.<sup>10</sup>

**2.2. Two empirical setups.** We focus on two particular empirical setups.

**Setup I: comprehensive treatment.** Consider a treatment that is comprehensive in the sense, that it is assigned and potentially administered to all eligible individuals at a common point in calendar time ("treatment day"), see figure 1a. The standard example here is a policy reform introduced via a change in legislation. A group of individuals who become unemployed at a common date is referred to as a cohort. In this setup,  $Z_i$  is the length of the time spell between the date of inflow into unemployment of individual  $i$  and the treatment day. When  $t < t'$ , we refer to the cohort  $\{Z = t\}$  (in short, the  $t$ -cohort) as the younger cohort and to the  $\{Z = t'\}$  cohort as the older one. If an individual  $i$  from the  $t$ -cohort remains unemployed until the treatment day, then  $S_i = t$  (enrollment into treatment) or  $S_i = \infty$ . If the individual finds a job (or, in general, exits the labor market)

---

<sup>10</sup> This distinction, however, is arbitrary and is mainly made for technical reasons, [Lancaster \(1990\)](#).

prior to the treatment day, then  $S_i$  is not observed. **Empirical example I.** In section 6, we

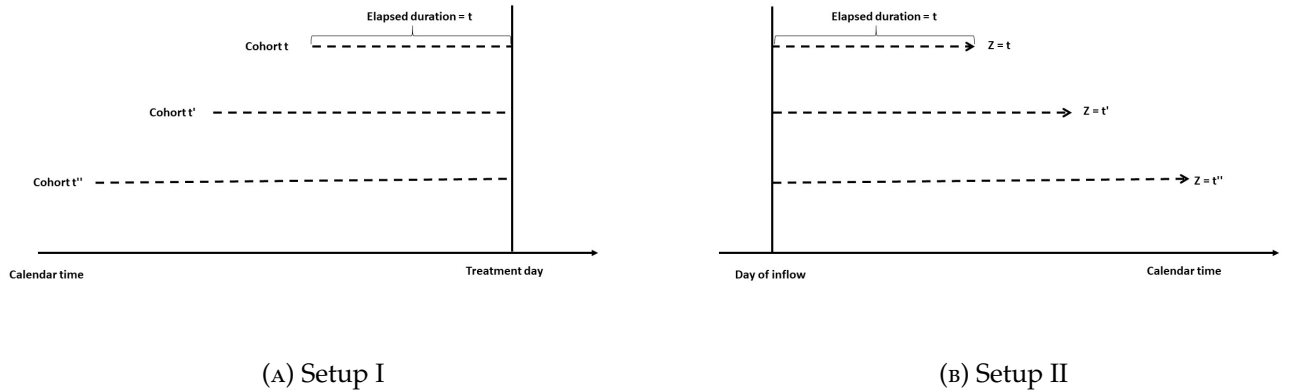


FIGURE 1. Two empirical setups

evaluate the French unemployment policy reform PARE which changed the payments structure of unemployment benefits and introduced ALMPs such as training. The new regime was effective from the 1st of July 2001 (comprehensive treatment assignment). Noncompliance was possible.

**Setup II: phased-in treatment.** Consider a group of individuals who enter the state of interest simultaneously (a single cohort). Treatment is assigned at different elapsed durations for different individuals, see figure 1b. **Empirical example II (a): time variation in ALMPs.** Often there is time variation in the administering of ALMPs across individuals, see e.g. [Lalive et al. \(2005\)](#), [Sianesi \(2004\)](#), [Crepon et al. \(2010\)](#) and [Abbring and van den Berg \(2003\)](#). Time variation might be necessary due to budgetary or other administrative reasons. If a case worker is responsible for a large group of unemployed, then meetings and coaching sessions must be assigned at different dates and hence at different elapsed durations of unemployment  $Z_i$ . **Empirical example II (b): phased-in implementation of social programs and field experiments.** Consider a social program that is characterized by a phased-in implementation: some units (regions or individuals) are assigned to be treated earlier, while others later. Phased-in implementation might be necessary for

similar reasons as in example II (a). Alternatively, a comprehensive program might be preceded by pilot studies introduced at varying times, which also gives rise to the setup described above. As an example, [Blundell et al. \(2004\)](#) utilize the presence of area-based pilot studies of the UK policy reform "New Deal for Young People" (NDYP) to identify its impact on unemployment.<sup>11</sup> An important subcategory is the phased-in (or pipeline, or rolled out) experimental design, in which the treatment is assigned at different times and the order is randomized. The seminal deworming study of [Miguel and Kremer \(2004\)](#) is the standard example here. Not-yet-treated individuals are taken as a control group for already treated individuals. Below, we explicitly formalize the assumptions that are typically made in this literature.

### 3. IDENTIFICATION OF DYNAMIC TREATMENT EFFECTS

3.1. **Assumptions.** Consider the following assumptions.

**Assumption A1 (Dynamic noncompliance).** *For any  $t$ , it holds either  $S(t) = t$  or  $S(t) = +\infty$ .*

Assumption A1 defines the possible type of noncompliance. Agents are only allowed to choose between being treated at the assigned point in time and being never treated. A1 precludes the type of choices  $S(t) = t'$  for some  $t' \neq t$  with  $t' < \infty$ . On the other hand, A1 allows individuals to change their preferences over time in an arbitrary way. As an example, for  $t < t'$ , individual  $i$  is allowed to choose  $S_i(t) = \infty$  and  $S_i(t') = t'$ . A noncomplier at  $t$  might be a complier at  $t'$ . In the context of an ALMP, an unemployed worker's decision whether to participate or not in an offered training at a given point in time will depend on the worker's subjective probability of the prospects to get a job without the training. Unsuccessful period of search might increase the readiness to participate as the individual gets more pessimistic.

---

<sup>11</sup> A further example for pilot studies is the implementation of the Progresa welfare reform in Mexico, see e.g. [Todd \(2007\)](#). Another example for phased-in social programs is the study of the effect of titling land properties on the labor market participation, [Field \(2007\)](#).

A1 mimics the standard static noncompliance model, in which the treatment is offered only once, see e.g. Heckman et al. (1999). At the same time, the process  $\{S(t)\}_{t \in \mathbb{R}^+ \cup \infty}$  extends the static concept of noncompliance by adding time structure. For a given point in time  $t$ , A1 corresponds to a one-sided noncompliance in the static treatment evaluation literature. One-sided noncompliance precludes the existence of always-takers and defiers.<sup>12</sup> As a result, no monotonicity-type assumption (as the one invoked in static LATE model) is needed for identification.

**Examples I and II (a), continued.** Assumption A1 is natural in the setup of the PARE reform, in which the treatment is administered at one single point in calendar time. Administrative and legislative rules prevent unemployed from enrolling into the treatment at earlier or later dates. In the context of the Swedish ALMPs, however, an assignment to treatment ( $Z$ ) by a case worker serves only as a recommendation. An unemployed is free to enroll into the treatment earlier or later, or never. Thus,  $S(t) = t'$  is possible for  $t' < t$  and  $t' > t$ , which violates A1.

**Assumption A2 (No anticipation).** Let  $\Theta_{T(s)}$  be the integrated hazard of  $T(s)$ . Then, for each real  $t' \geq t \geq 0$ , it holds

$$(3.1) \quad \Theta_{T(t')}(t \mid X, V, S(t), S(t')) = \Theta_{T(\infty)}(t \mid X, V, S(t), S(t')).$$

A2 states that present potential outcomes are not influenced by future events. This can be more clearly the following relationship

$$(3.2) \quad P(T(t') > t \mid X, V, S(t), S(t')) = P(T(\infty) > t \mid X, V, S(t), S(t')).$$

Equality (3.2) is equivalent to A2. It states that the individual survival probability up to some earlier time  $t$  remains the same under any potential future treatment times  $t', t''$  ( $t \leq t', t''$ ). Conditioning on  $S(t), S(t')$  implies that A2 is valid for the subgroups of

---

<sup>12</sup> To see this, note that A1 precludes choices  $S(\infty) = t$  for  $t < \infty$ . See Imbens and Angrist (1994) for the definitions of always-takers and never-takers.

compliers and noncompliers at  $t$  and  $t'$ . In a setting with forward-looking individuals, A2 is satisfied when the information structure is invariant to the potential assignment of the treatment. There are two major cases when A2 can be viewed as plausible. The first case is when individuals have no knowledge on the point in time of treatment (i.e. they do not anticipate it). As an example, the assignment to ALMPs may occur without a preliminary notice so that the timing is unexpected to the unemployed. This is almost by definition true for some punitive treatments such as sanctions. The second case is when individuals do not act upon the knowledge of the time to treatment. As an example, the treatment might be so complex or the consequences so ambiguous, that the resulting uncertainty deters the unemployed from adapting their behavior. In the context of the PARE reform (example I), the unemployed individuals were informed on a short notice (2 weeks) about the upcoming reform. The exact content and the start of the reform were subject to persistent political debate, so that its actual implementation came as a surprise.

*No anticipation in Setup II.* A2 is implicitly assumed in all experiments with a phased-in design (example II (b)). A unit treated at  $t'$  can be used at  $t, t < t'$ , as a control for a unit treated at  $t$  only if the not-yet-treated unit does not anticipate the treatment at  $t'$ . Violation of this assumption is considered to be one major potential flaw in the evaluation of phased-in treatments, [Duflo et al. \(2007\)](#). Note that phased-in experiments are typically evaluated in a static framework. The “no anticipation” assumption is hidden in the randomization assumption. In a dynamic framework, however, (initial) randomization is not sufficient since evaluation typically conditions upon survival. Due to dynamic selection, the composition in the different treatment arms might change differently. Here, A2 is sufficient to complement a static randomization assumption. We establish the link explicitly in the next section.

*No anticipation in Setup I.* A2 has a subtle additional implication in Setup I. Because the spells of treatment and control cohorts begin at different calendar dates, an elapsed



duration of  $t$  time units is also reached at different dates. Thus, expression (3.1) not only requires that individuals do not anticipate the (date of the) treatment, but also that the economic conditions of treatment and control cohorts are identical. This requirement can be seen as a stationarity requirement on the data generation process (no cohort effects). As an example, if the local labor market conditions change substantially between two cohorts, then assumption A2 would be violated even if individuals do not anticipate the treatment.

*No anticipation in the dynamic literature.* In the context of ALMPs, the “no anticipation” assumption has been adopted throughout the theoretical and empirical literature on dynamic treatment effects, see e.g. [Sianesi \(2004\)](#), [Vikström \(2014\)](#), [Abbring and van den Berg \(2003\)](#). Much of this literature is surveyed in [Crepon et al. \(2010\)](#).

*No anticipation in the static evaluation literature.* Assumption A2 is often implicitly assumed in static evaluation approaches. One example is the phased-in experimental design discussed above. Further example is the DiD approach. Anticipation effects potentially undermine the parallel trends assumption, see e.g. [Lalive \(2008\)](#) for an application in a labor market context. A third example is the Synthetic Control approach, where A2 is an implicit component of the conditional independence assumption.

**Assumption A3 (Randomization).** *It holds*

$$(3.3) \quad (i) \quad Z \perp\!\!\!\perp \{T(s), S(t)\}_{t,s \in \mathbb{R}_+ \cup \{+\infty\}} \mid X, V \quad \text{and} \quad (ii) \quad Z \perp\!\!\!\perp V \mid X.$$

Assumption A3 is a randomization assumption. In the context of Setup II, assumption A3 (i) postulates that assignment to treatment ( $Z$ ) is independent of potential outcomes conditional on observed and unobserved covariates. Since  $X$  and  $V$  are assumed to fully describe an individual, this is an innocuous assumption. The major component of A3 is assumption A3 (ii). It states that the instrument  $Z$  is independent of unobservables

conditionally on observables. It holds the following implication:

$$(3.4) \quad A3(i), A3(ii) \Rightarrow Z \perp \{T(s), S(t)\} \mid X.$$

Thus, under A3, assignment of  $Z$  is driven by  $X$ . **Example II (a), continued.** Suppose that a case worker is responsible for a large pool of unemployed individuals. Then it can be argued that she acts idiosyncratically given (objective) characteristics of the unemployed and her own assessment of the unemployed.<sup>13</sup> In such cases, A3 is valid if both the objective characteristics and the case worker's assessment are also available to the econometrician. As a further example consider randomized phased-in experiments. A3 is valid per design construction. In the case of phased-in implementation of social programs, A3 holds whenever the early/late recipients are randomly selected. In Setup I, the plausibility of A3 hinges on the stability of the economic environment: A3 requires identical economic environments at the dates of inflow of young and old cohorts. The relation to A3 in Setup II is best explained with a thought experiment, which features individuals randomly assigned to different cohorts. Differences in the structural economic parameters of treated vs nontreated cohorts (cohort effects) violate A3. Such differences can be caused e.g. by mass layoffs and macroeconomic trends.

**Assumption A4 (Consistency).** For all  $t, s \in \mathbb{R}_+ \cup \{+\infty\}$

$$i) \quad Z = t \Rightarrow S(t) = S$$

$$ii) \quad S = s \Rightarrow T(s) = T.$$

The consistency assumption states that a potential outcome corresponding to a given treatment is observed if the treatment is actually assigned. Another way to write it is  $T = T(S), S = S(Z)$ . This is a standard assumption in the treatment evaluation literature. It

---

<sup>13</sup> As an example, in the paper of [Sianesi \(2004\)](#), such assessments relate to the job-seeker's degree of job readiness, as well as to the job-seeker's inclinations and urgency to find a job. These assessments are documented and part of the observed covariates used in [Sianesi \(2004\)](#).

provides the link between potential outcomes and observables. Assumptions A1 and A4 imply together that the actual elapsed duration at which the treatment is received,  $S$ , can be either equal to  $Z$  or to  $\infty$ .

In addition to assumptions A1-A4, we implicitly assume that all expressions below exist. This amounts to common support assumptions such as  $0 < P(S = t \mid X, V, Z = t)$ . These assumptions are fulfilled for example when  $S$  and  $Z$  are discrete, but it is sufficient that  $S$  and  $Z$  have a positive probability mass on  $t$  and  $t'$ . Whether discrete  $Z$  and  $S$  impose a restriction on the distribution of  $T$  depends on the concrete application. In the medical treatment example, a specific therapy might be assigned only at predetermined, common for everybody, elapsed time intervals of the disease, whereas the life or disease duration itself is a continuous variable. In the labor market example, the administrative duration of unemployment is always discrete. Nevertheless, it is usually modeled in the literature as a continuous variable, especially when it is measured on a daily basis. On the other hand, labor market treatments such as training and counseling measures or financial penalties might be designed to come into force only at coarser time intervals. Therefore, it might be practical to model them as discrete variables.

**Remark 6.** (*Potential biases in Setup I.*)<sup>14</sup> Denote the calendar date of treatment in Setup I with day 0. Consider cohorts  $Z = t$  and  $Z = t'$  which enter unemployment at dates  $-t$  and  $-t'$ , respectively, with  $t < t'$ . From the discussion of A2 and A3, nonstationary economic environment emerges as a possible source of bias in Setup I. We distinguish between two cases. Case 1: the economic environment is stationary up to date 0, and there is a structural change after 0. This can be the case, for example, when other economic policies are implemented at 0 alongside with the ALMP of interest. One approach to isolate the effect of the ALMP of interest would be to consider the limit case  $a \rightarrow 0$ . The validity of this

---

<sup>14</sup> This remark and the resulting RD analysis in section 3.3 below result from a hint by an anonymous referee, for which we are thankful.

approach would hinge upon the assumption that the change in the economic environment at/after 0 has no instantaneous effect on the unemployment duration. In this case, our main identification strategy described in section 3.2 below would yield an unbiased estimator of the treatment effect on the hazard. Case 2: in the general case of different economic conditions, the hazard approach is not sufficient. In this case, one approach would be to consider the limit case  $t' \rightarrow t+$ , where the + sign indicates convergence from the right. The intuition of this approach is that individuals that enter unemployment at two points in time that are very close are basically evaluated in the same economic environment. This strategy will lead to RD-type estimator, see section 3.3. A drawback of both approaches (Case 1 and Case 2) is the limitation of the set of treatment effects we can evaluate. In both cases, namely, we restrict the window  $a$  to be very short.<sup>15</sup> This drawback is closely related to the critique of the phased-in design experiments, in which evaluation of long-term effects is in general not possible, [Duflo et al. \(2007\)](#).

**3.2. Identification results.** Assume first that  $T$  is observable. Consider expression (2.2). The main challenge is to find a control group for those who survived until  $t$  and were treated at  $t$ . In particular,  $T(t')$  and  $S(t)$  are never jointly observed. They correspond to different treatments ( $t'$  and  $t$ ). Thus, one of the outcomes  $T(t'), S(t)$  is always counterfactual. To motivate our identification strategy, consider the following naive candidates for a treatment effect:

$$(3.5) \quad P(T \in [t, t+a) \mid T \geq t, X, S = t, Z = t) - P(T \in [t, t+a) \mid T \geq t, X, S = \infty, Z = t),$$

$$(3.6) \quad P(T \in [t, t+a) \mid T \geq t, X, S = t, Z = t) - P(T \in [t, t+a) \mid T \geq t, X, S = t', Z = t')$$

---

<sup>15</sup> In the second case,  $a \rightarrow 0$  is implied by  $a \leq t' - t$ .

for  $t' > t$ . For simplicity, we set the discussion in the context of Setup I. Writing (3.5) in the form

$$\begin{aligned} & \mathbb{E}_{V|T \geq t, X, S=t, Z=t} [P(T \in [t, t+a) | T \geq t, X, S=t, Z=t, V)] - \\ & \mathbb{E}_{V|T \geq t, X, S=\infty, Z=t} [P(T \in [t, t+a) | T \geq t, X, S=\infty, Z=t, V)] \end{aligned}$$

makes it clear that (3.5) compares averages over two different subgroups of the same cohort: the  $t$ -compliers and the  $t$ -noncompliers. Since  $S$  is a choice variable, it holds in general

$$(3.7) \quad V \not\perp S | T \geq t, X, Z = t.$$

Thus, (3.5) captures not only the treatment effect but also the selection bias. The non-surprising implication is that noncompliers are not suitable as a control group.

Expression (3.6) compares the average outcome at elapsed duration  $t$  of the  $t$ -compliers from the younger cohort  $\{Z = t\}$  with the average outcome (at the same duration) of the  $t'$ -compliers from the older cohort  $\{Z = t'\}$ . In general, however,  $F_{V|T \geq t, S=t, Z=t} \neq F_{V|T \geq t, S=t', Z=t'}$  due to dynamic selection. This follows because some unemployed might find a job between elapsed durations  $t$  and  $t'$ , while others might change their preferences. As a result, learning the compliance status at  $t'$  is also not helpful for constructing a control group. The above considerations apply equivalently for Setup II.

Instead, our strategy combines the approach of the dynamic treatment effects literature with the static LATE approach. Thus, identification consists of two steps: a dynamic and a static one. Our dynamic step, presented in the next proposition, extends the result of [Van den Berg et al. \(2014\)](#) to a setting with endogenous noncompliance.<sup>16</sup>

---

<sup>16</sup> All proofs are in the online appendix.

**Proposition 3.1.** *Let  $F$  be a cdf. Under assumptions A1 to A4, it holds for all  $\infty \geq t' \geq t \geq 0$*

$$(3.8) \quad F_{V|T(t) \geq t, X, S(t)=t} = F_{V|T(t') \geq t, X, S(t)=t} = F_{V|T \geq t, X, S=t, Z=t} \quad \text{and}$$

$$(3.9) \quad F_{V|T(t) \geq t, X, S(t)=\infty} = F_{V|T(t') \geq t, X, S(t)=\infty} = F_{V|T \geq t, X, S=\infty, Z=t}.$$

To interpret proposition 3.1, consider example I. Assume that we observe two cohorts ( $t$  and  $t'$ ) of unemployed individuals with dates of inflow  $-t = -6$  and  $-t' = -9$  (0 is set to be equal to 01.07.2001, the day of the implementation of the reform). The first equality in (3.8) states that under A1 - A4, individuals in cohort  $t$  who have been unemployed for at least 6 months and are willing to take the treatment have the same distribution of  $V$  as individuals in cohort  $t'$  who also have been unemployed for at least 6 months and are willing to take the treatment. The second equality of (3.8) links potential to observed conditions. (3.9) provides an equivalent result for the group of  $t$ -noncompliers. Note, that an implication of proposition 3.1 is that the treatment effects on the treated survivors,  $\{T(t) \geq t\}$ , and on the nontreated survivors,  $\{T(t') \geq t\}$ , and hence on all survivors, coincide.

The intuition behind proposition 3.1 is the following. If the unobserved heterogeneity  $V$  has the same distribution in the two cohorts at the point in time of inflow (conditional on  $X$ ), and if these distributions evolve over time in the same way, then  $V$  will have the same distribution in the two cohorts at a later pre-treatment elapsed duration  $t > 0$ , see the dotted lines in figures 2a and 2b. The equality of the distributions of  $V$  at  $t = 0$  is ensured by the randomization assumptions A3. The dynamics is controlled by the "no anticipation" assumption A2. The interpretation and intuition for Setup II are equivalent with cohorts replaced by groups of individuals who are assigned to the treatment at different dates.

To motivate the second step of our approach (the static step), consider the following corollary of proposition 3.1.

**Corollary 3.1.** *Let  $a \leq t' - t$ . Under Assumptions A1-A4, it holds for all  $\infty \geq t' \geq t \geq 0$*

$$(3.10) \quad \begin{aligned} TE(t, t', a) &= P(T(t) \in [t, t + a) \mid T(t) \geq t, X, S(t) = t) \\ &\quad - P(T(t') \in [t, t + a) \mid T(t') \geq t, X, S(t) = t). \end{aligned}$$

The r.h.s. of (3.10) does not contain the unobserved  $V$ . Thus, if the compliance status ( $S(t)$ ) was observed, the treatment effect could be calculated by comparing the average outcome of compliers from the treated group  $\{Z = t\}$ , denoted by  $F_{1,C}$ , with the average outcome of compliers from the not-yet-treated group,  $\{Z = t'\}$ , denoted by  $F_{0,C}$ . For the subgroup  $\{Z = t\}$ ,  $S(t)$  is observed at elapsed duration  $t$ . Therefore,  $F_{1,C}$  is identified. For the not-yet-treated  $\{Z = t'\}$ , however,  $S(t)$  is unobserved.

Our static step is as follows. Consider the average potential outcome of all nontreated survivors,  $F_0 = P(T(t') \in [t, t + a) \mid T(t') \geq t)$ , where the dependency on  $X$  is suppressed. The key to identification is the decomposition of  $F_0$

$$(3.11) \quad F_0 = F_{0,C}P_{0,C} + F_{0,N}P_{0,N},$$

where  $F_{0,N}$  is the average outcome for noncompliers in the nontreated group, and  $P_{0,C}$  and  $P_{0,N}$  are the proportions of compliers and noncompliers in that group, respectively (all at elapsed duration  $t$ ). Figures 2a and 2b illustrate the idea. Solving (3.11) for  $F_{0,C}$  yields  $F_{0,C} = (F_0 - F_{0,N}P_{0,N})/P_{0,C}$ . Hence, in order to identify  $F_{0,C}$  it is sufficient to identify  $F_0, P_{0,C}, P_{0,N}$  and  $F_{0,N}$ . We can directly link  $F_0$  to observable outcomes at elapsed duration  $t$ : it is equal to the average observed outcome at  $t$  of the whole not-yet-treated group,  $P(T \in [t, t + a) \mid T \geq t, Z = t')$ . Thus,  $F_0$  is identified. In addition, under assumptions A1 – A4, we can identify  $P_{0,C}, P_{0,N}$  and  $F_{0,N}$  from the treated group. In particular, due to randomization and no-anticipation, all pre-treatment characteristics of the two groups have equal distributions at elapsed duration  $t$ . It holds therefore  $P_{0,C} = P_{1,C}, P_{0,N} = P_{1,N}$  and  $F_{0,N} = F_{1,N}$ . The intuition for the last equality is that in both groups noncompliers are

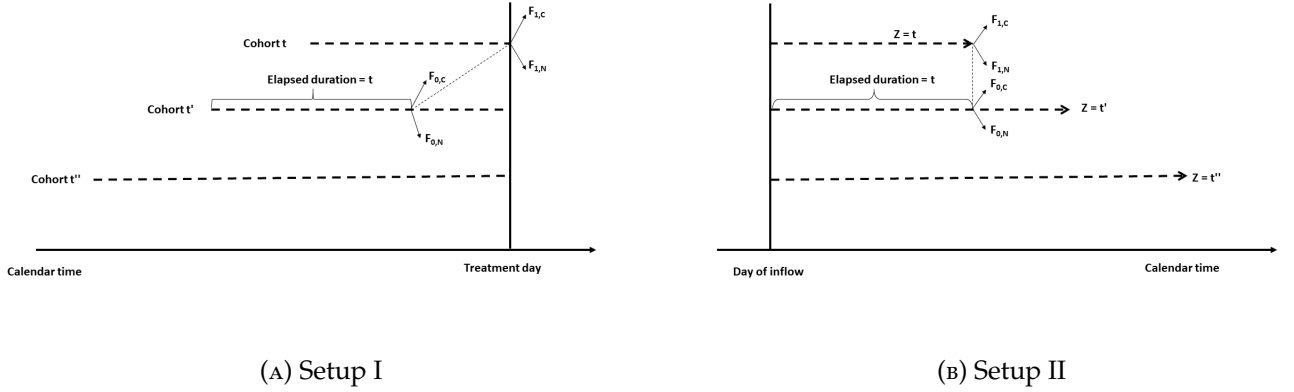


FIGURE 2. Identification in the two setups

not treated. The exclusion restriction (2.1) ensures that the assignment to treatment alone does not change their outcomes.

With these considerations we can state our main identification result:

**Proposition 3.2.** *Let  $a \leq t' - t$ . Under Assumptions A1-A4, the treatment effect on the treated  $TE(t, t', a)$  is nonparametrically identified for all  $\infty \geq t' \geq t \geq 0$  and it holds*

$$(3.12) \quad TE(t, t', a) = \frac{P(T \in [t, t+a] \mid T \geq t, X, Z = t) - P(T \in [t, t+a] \mid T \geq t, X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)}$$

Expression (3.12) has an intuitive interpretation. It adjusts the difference between the average observed outcomes in the two groups by the probability to be a complier. The adjustment takes account of the fact, that any difference between the outcomes of the two groups can be caused only by the compliers. Our result is related to the static one-sided noncompliance result of [Bloom \(1984\)](#) and the LATE identification result in [Imbens and Angrist \(1994\)](#). Identification is local in the sense, that the treatment effect is identified only for the subgroup of  $t$ -compliers. As this group is allowed to change with  $t$ , our notion of location can be seen as a dynamic extension of the LATE notion of location.

We now consider the case of right censoring. In labor market context, right censoring typically arises when at the end of the period of observation some individuals are still unemployed, so their unemployment spells have an unknown length. Censoring occurs



also when unemployment is interrupted by a transition out of the labor force due to maternity, sickness and military service, or simply when individuals do not show up to report about their status (attrition). With a small abuse of notation, let  $C \geq 0$  be the censoring r.v.. Define  $\tilde{T} = \min\{T, C\}$  and  $\delta = \mathbb{1}\{\tilde{T} = T\}$ . We observe  $(\tilde{T}, \delta)$  and not directly  $(T, C)$ . It is not possible to recover nonparametrically the joint distribution of  $T$  and  $C$  from the distribution of  $(\tilde{T}, \delta)$  without additional assumptions, see [Tsiatis \(1975\)](#). We adopt the following additional standard assumption:

**Assumption A5 (Random censoring).**  $C \perp (T, S) \mid X, Z$ .

It holds the following proposition.

**Proposition 3.3.** *Under assumptions A1 - A5  $TE(t, a)$  is identified.*

**Remarks.7.** Identification of a treatment effect on the hazard (*HTE*) requires only mild additional regularity assumption and is left for the appendix, section . 8. Under A1-A4, we have  $P(T \in [t, t + a) \mid T \geq t, X, Z = t') = P(T \in [t, t + a) \mid T \geq t, X, Z = t'')$  for all  $t', t'' \geq t + a$  (in the limit case  $a \rightarrow 0$  simply for  $t', t'' > t$ ). As a consequence, the treatment effect does not depend on the choice of the nontreated cohort  $t'$  as long as  $t' \geq t + a$  (or  $t' > t$ ). Therefore, we omit the dependence on  $t'$  and write  $TE(t, a)$  and  $HTE(t)$ .

**3.3. Identification in a Regression Discontinuity setup.** In this section, we outline an identification approach that mitigates the problems related to Setup I discussed in remark 6.<sup>17</sup> The intuition is that if  $t'$  is sufficiently close to  $t$ ,  $t < t'$ , then treatment and control cohorts are practically evaluated in the same economic environment. This consideration leads to the following modifications of A2 and A3:

**Assumption A2':**  $P(T(t) > t \mid X) = \lim_{t' \rightarrow t+} P(T(t') > t \mid X)$ ,

**Assumption A3':** Denote by  $F_{T(s), S(z) \mid Z, X, V}$  and  $F_{V \mid Z, X}$  the corresponding conditional distributions. Then there exist a positive number  $\eta$ , such that for all  $t'$  in the  $\eta$ -neighbourhood

<sup>17</sup> We are thankful to an anonymous referee for suggesting this strategy to us.

of  $t$  and for all positive  $(s, z)$

$$(i) F_{T(s), S(z) | Z=t, X, V} = F_{T(s), S(z) | Z=t', X, V} \quad \text{and} \quad (ii) F_{V | Z=t, X} = F_{V | Z=t', X}.$$

Assumptions A2' and A3' are local versions of A2 and A3, respectively. Note that A2' is almost trivially fulfilled in both empirical Setups. In particular, even if an individual has perfect knowledge of the assigned treatment ( $t$  or  $t'$ ), the effect of this knowledge under  $t$  and  $t'$  will be the same when  $t' \rightarrow t+$ .

Let  $\theta(\cdot | X, Z)$  be the conditional hazard function of  $T$  and assume the mild regularity condition HTE1 (see section A in the appendix for a formal definition of the hazard and for assumption HTE1). The following proposition can now be stated.

**Proposition 3.4.** *Assume that  $\lim_{t' \rightarrow t+} \theta(t | X, Z = t')$  exists. Then, under assumptions A1, A2', A3', A4, A5 and HTE1, the treatment effect on the hazard at  $t$  is identified and equal to*

$$(3.13) \quad \Psi_{RD}(t) = \lim_{t' \rightarrow t+} \frac{\theta(t | X, Z = t) - \theta(t | X, Z = t')}{P\{S = t | T \geq t, X, Z = t\}}.$$

(3.13) is related to the static Regression Discontinuity (RD) identification result in [Hahn et al. \(2001\)](#).  $Z$  can be seen as a forcing variable with a discontinuity is at  $t$ . Assumption A3' is related to the conditional independence assumption made in Theorem 2 in [Hahn et al. \(2001\)](#). Assumption A2', however, is new and needed to account for dynamic selection. Thus, an estimator of the treatment effect based on A2', A3' and result (3.13) can be interpreted as a dynamic RD estimator.

We do not implement this approach in our empirical application due to a practical problem. In particular, the finite sample performance of the estimator might lack precision due to lack of observations near the boundary. It is out of the scope of this paper to deal with this problem. In addition, note that if the treatment effect does not substantiate instantaneously, it will not be detected by the estimator.

## 4. ESTIMATION

In this section, we develop an estimator for  $TE(t, a)$ . An estimator for the treatment effect on the hazard is presented in section A.2 in the appendix. For simplicity of exposition, we ignore the dependence on covariates. All results below generalize in a straightforward way to the case with covariates. One simply uses the conditional Kaplan-Meier estimator of [Gonzalez-Manteiga and Cadarso-Suarez \(2007\)](#) instead of its unconditional counterpart.

Let  $t < t'$ . Define  $\bar{F}_1(t) = P\{T > t \mid Z = t\}$ ,  $\bar{F}_2(t) = P\{T > t \mid Z = t'\}$  and  $p = P\{S = t \mid T \geq t, Z = t\}$ . The former two are observed survival probabilities and the latter is the compliance probability. Under A1-A4,  $TE(t, a)$  can be written as

$$(4.1) \quad TE(t, a) = \frac{1}{p} \left( \frac{\bar{F}_2(t+a)}{\bar{F}_2(t)} - \frac{\bar{F}_1(t+a)}{\bar{F}_1(t)} \right).$$

Furthermore, we allow  $T$  to be right censored, and we assume that we have access to i.i.d. observations  $(\tilde{T}_i, \delta_i, S_i, Z_i)$ ,  $i = 1, \dots, n$ . Denote by  $\widehat{F}_j$  the nonparametric Kaplan-Meier estimator of  $\bar{F}_j$ ,  $j = 1, 2$ . Consider the following high level assumptions:

$$(4.2) \quad \widehat{F}_j(t) = \bar{F}_j(t) + o_p(1),$$

$$(4.3) \quad \sqrt{n} \left( \widehat{F}_j(t) - \bar{F}_j(t) \right) \xrightarrow{d} N(0, \sigma_j^2(t)) \quad \text{as } n \rightarrow \infty,$$

where  $\sigma_j^2(t)$  is the asymptotic variance of the Kaplan-Meier estimator,  $t \in [0, \infty)$ , see e.g. page 18 ff. [Kalbfleisch and Prentice \(2002\)](#). These conditions follow from mild regularity conditions that can be found in standard references for survival analysis, see e.g. [Andersen et al. \(1997\)](#), chapter IV.3 or [Kalbfleisch and Prentice \(2002\)](#), chapter 5.6. We do not state them explicitly. Finally, let  $\widehat{p}$  be a consistent nonparametric estimator of  $p$ . With those preliminaries, we define the IV-estimator  $\widehat{TE}(t, a)$  of  $TE(t, a)$  as

$$(4.4) \quad \widehat{TE}(t, a) = \frac{1}{\widehat{p}} \left( \frac{\widehat{F}_2(t+a)}{\widehat{F}_2(t)} - \frac{\widehat{F}_1(t+a)}{\widehat{F}_1(t)} \right).$$

(4.4) can be interpreted as a dynamic version of a Wald estimator. Its consistency is stated in the following proposition.

**Proposition 4.1.** *Suppose (4.2) holds. Then, under assumptions A1-A5, it holds*

$$\widehat{TE}(t, a) - TE(t, a) = o_p(1).$$

The following proposition states the asymptotic distribution of the estimator.

**Proposition 4.2.** *Let assumptions A1-A5 and condition (4.3) hold. Then it holds*

(4.5)

$$\sqrt{n}(\widehat{TE}(t, a) - TE(t, a)) \xrightarrow{d} N\left(0, \frac{1}{p^2} \sum_{i=1}^2 \left( \frac{1}{\bar{F}_i^2(t)} \sigma_i(t+a) + \frac{\bar{F}_i^2(t+a)}{\bar{F}_i^4(t)} \sigma_i(t) + \frac{\bar{F}_i(t+a)}{\bar{F}_i^3(t)} \sigma_i(t, t+a) \right)\right),$$

where  $\sigma_i(t, t+a)$  is the covariance of  $\widehat{F}_i(t)$  and  $\widehat{F}_i(t+a)$ .

Confidence bands can be constructed by replacing the unknown terms in the variance with consistent estimates, for example using the Greenwood's formula, see [Andersen et al. \(1997\)](#). It follows from (4.5) that the precision of the estimator is inversely related to  $p$ . The bigger the compliance probability  $p$  (i.e. the stronger the instrument  $Z$ ), the smaller the variance.

## 5. MODEL DIAGNOSTICS

**5.1. Testing the assumptions.** We suggest two testing approaches to address the main assumptions A2 and A3.

**Pre-treatment approach.** The first approach parallels model diagnostics in a DiD context and focuses on pre-treatment outcomes. Empirical tests for equality of pre-treatment survival probabilities are applied by [De Giorgi \(2005\)](#) and [Van den Berg et al. \(2014\)](#). In this section, we give a theoretical justification for these tests. Let  $s \leq t < t'$ . Under A1, A4, assumptions A2 and A3 jointly imply

$$(5.1) \quad P(T > s \mid X, Z = t) = P(T > s \mid X, Z = t').$$

In words, observed survival probabilities are equal in the treated and nontreated groups at any common elapsed pre-treatment duration. Adopting A1 and A4 as fundamental assumptions, equality (5.1) can be used to test for no anticipation and randomization. This is captured in the following proposition (dependency on  $X$  is ignored).

**Proposition 5.1.** *Let assumptions A1, A4, A5 and condition (4.3) hold. Then, under the null hypothesis  $H_0 : A2 \cup A3$ , it holds*

$$(5.2) \quad \sqrt{N}(\widehat{F}_1(t) - \widehat{F}_2(t)) \stackrel{d}{\approx} \mathbf{N}(0, 2\sigma^2(t)).$$

**RD approach** The second approach is to adopt the weaker assumptions A2' and A3' (together with A1, A4, and A5) as fundamental and use (3.13) in a DiD setup. Let  $s > t$ . Then under A2 and A3, the expression

$$(5.3) \quad (\theta(t | X, Z = t) - \theta(t | X, Z = s)) - (\theta(t | X, Z = t) - \lim_{t' \rightarrow t+} \theta(t | X, Z = t')) = \lim_{t' \rightarrow t+} \theta(t | X, Z = t') - \theta(t | X, Z = s)$$

is equal to 0. A test statistic can be constructed along the lines of section A.2 in the appendix.

**5.2. Framework for the analysis of endogeneity.** A comprehensive policy reform is often preceded by a small scale pilot study that allows for noncompliance, see e.g. [Todd \(2007\)](#). Understanding the non-take up of the pilot study might help better design the reform and derive bounds for its effect under perfect compliance. The motivation of our analysis is that the individuals might select into treatment based on potential outcomes, [Heckman \(2008\)](#). Therefore, one approach to analyze noncompliance is to test for equality of potential outcomes. In particular, for some  $t < t'$ , we are interested in testing

$$(5.4) \quad (i) F_{0,C} = F_{0,N} \quad \text{and} \quad (ii) F_{1,C} = F_{1,N},$$

where we used the simplified notation from section 3.2. It follows from assumption A1 that  $F_{1,N}$  is not identified:  $t$ -noncompliers are never observed under the treatment  $t$ . As

a result, hypothesis (5.4) (ii) is not testable. We therefore focus on (5.4) (i). It follows from relation (3.11) that hypothesis (5.4) (i) is equivalent to  $F_0 = F_{0,N}$ . In particular, if compliers and noncompliers have equal (average) potential outcome distributions, then the outcome distribution of the whole population under no treatment is equal to the outcome distribution of the noncompliers under no treatment. Furthermore, we showed in section 3.2 that  $F_0$  and  $F_{0,N}$  are identified under A1-A4, with

$$F_0 = P(T \in [t, t+a] \mid T \geq t, X, Z = t'),$$

$$F_{0,N} = F_{1,N} = P(T \in [t, t+a] \mid T \geq t, X, S = \infty, Z = t).$$

These considerations lead to the testable hypothesis  $\mathbb{D} = 0$ , where

$$\mathbb{D} = P(T \in [t, t+a] \mid T \geq t, X, Z = t') - P(T \in [t, t+a] \mid T \geq t, X, S = \infty, Z = t).$$

$\mathbb{D}$  is the difference of the observed outcome distributions of the not-yet-treated and the noncompliers from the treated group. The test statistic is constructed along the lines of section 4.<sup>18</sup>

The bias that arises from an endogenous non-take up can be measured as the difference between the true treatment effect (3.12),  $TE = F_{1,C} - F_{0,C}$ , and the naive candidate (3.5) (for short  $NTE$ ),  $NTE = F_{1,C} - F_{1,N}$  (dependence on  $t$  and  $a$  is suppressed). Define  $\mathbb{B} = TE - NTE$ . Substituting  $F_{0,C} = (F_0 - F_{0,N}P_N)/P_C$  and  $F_{0,N} = F_{1,N}$  yields  $\mathbb{B} = (F_{0,N} - F_0)/P_C = \mathbb{D}/P_C$ . An empirical analysis of the bias from endogenous selection can be performed with an estimator of  $\mathbb{B}$ .

<sup>18</sup> A simplified testing procedure would induce a comparison of unconditional survival functions. The corresponding null hypothesis is

$$(5.5) \quad \tilde{H}_0 : \quad P(T \geq t \mid X, Z = t') - P(T \geq t \mid X, S = \infty, Z = t) = 0.$$

## 6. EMPIRICAL APPLICATION: THE FRENCH PARE LABOR MARKET REFORM IN 2001

In this section, we illustrate our methods in the context of the French labor market reform Plan d'Aide au Retour à l'Emploi, PARE for short (example I).<sup>19</sup> Under the old system (prior July 1st 2001), the amount of individual unemployment benefits (UB) is a stepwise decreasing function of time. Under the new regime (post July 1st 2001), the UB are constant over the whole payment period of an eligible individual. In addition, the reform introduces a range of ALMPs such as compulsory meetings with a case worker and job search training.<sup>20</sup> The more generous UB rules and the new ALMPs have potentially opposite effects on the duration of unemployment. Thus, ex ante it is not clear what the overall effect of the policy will be.

A distinct feature of the reform is that interrupted spells - i.e. those individuals who became unemployed prior to July 1st and were still unemployed on that day - could choose between the old and the new regulations. New spells (whose inflow is after July 1st 2001) were automatically assigned to the new system.

We evaluate the effect of the reform on the duration of unemployment for those who have been unemployed for at least  $t = 6$  months.<sup>21</sup> Thus, the theoretical treatment effect of interest is defined as

$$(6.1) \quad TE(6, a) = P(T(6) \in [6, 6 + a) \mid T(6) \geq 6) - P(T(t') \in [6, 6 + a) \mid T(t') \geq 6),$$

where  $t' > 6$  and  $a < t' - 6$ .<sup>22</sup>

<sup>19</sup> Some studies that evaluate the French unemployment insurance system are e.g. [Fougère et al. \(2010\)](#), [Debauche and Jugnot \(2007\)](#), [Crépon et al. \(2012\)](#), [Crépon et al. \(2005\)](#) and [Le Barbanchon \(2012\)](#).

<sup>20</sup> A comprehensive description of the reform can be found in [Freyssinet \(2002\)](#).

<sup>21</sup> We describe the dataset and our empirical strategy (and in particular the motivation behind our choice of the treated cohort) in section B.1 in the appendix.

<sup>22</sup> We use here corollary 3.1 to simplify the expression for the treatment effect.

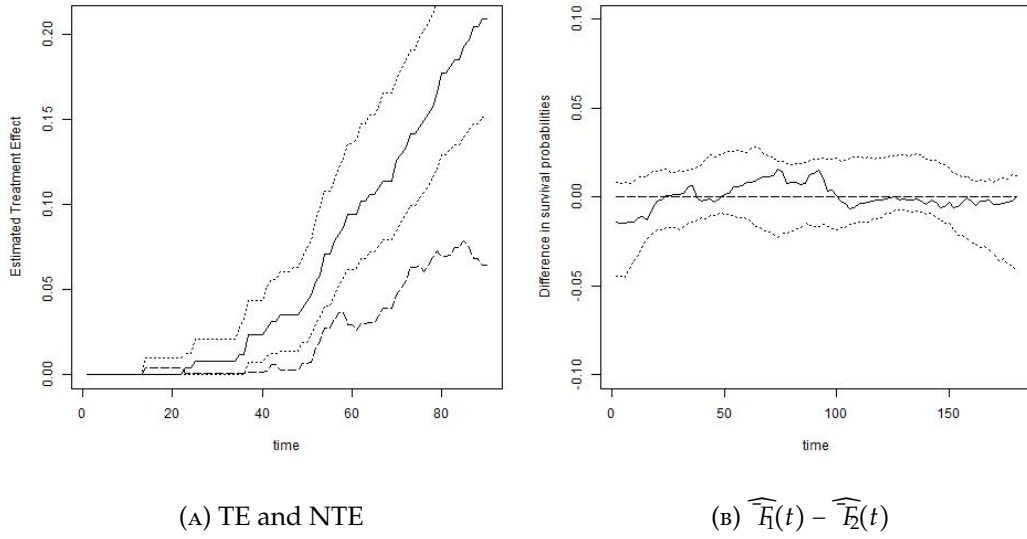


FIGURE 3. (a): Estimated treatment effect (thick line), 95% confidence bounds (dotted line), Naive treatment effect (dashed line). (b): Difference of pre-treatment survival functions (thick line), 95% confidence bounds (dotted line), the zero axis (dashed line).

The estimated treatment effect is presented in figure 3a for the choice  $t' = 9$  months of the control group and different values of  $a \in [0, 90)$  ( $t' - t = 3$  months = 90 days). The estimate  $\widehat{TE}(6, a)$  is represented by a thick line, and the 95 %-confidence bounds are represented by dotted lines. The effect is positive and increasing with  $a$ , and for  $a \geq 37$  it is also significant. According to our estimates, the probability to find a job within the first three months after receiving the treatment increased with up to 0.2 compared to the counterfactual case where the treated would not have received the treatment. In section B.2 in the appendix, we present results for different subgroups in order to analyze potential treatment effect heterogeneity. The estimates for the subgroups of white and blue collar workers, and for the subgroups of unemployed with higher and lower education status follow a pattern very similar to the pattern of the unconditional estimates.

Next, we perform thorough model diagnostics. First, chi-square tests for equality of distributions reveal that pre-treatment observed characteristics are balanced between the treatment and control groups, see appendix B.3.1. Second, we also provide evidence that



the economic condition at the inflow of the two cohorts were very similar, also in appendix B.3.1. These findings support the plausibility of A3, and in particular that the estimates are not undermined by cohort effects. Third, we estimate the TE with an alternative choice of the control cohort ( $t'' = 8$  months). The estimation results are very similar to the main results, see figure 6 in the appendix. This is further evidence that there were no cohort effects. In section B.3.2, we provide arguments that the reform was not anticipated to the unemployed due to the relative short notice and lack of (clear) political debate. Fourth, we test for equality of pre-treatment survival functions along the lines of section 5.1. The estimated difference of the survival functions of treatment and control groups is depicted in figure 3b with a thick line. The zero axis lies is fully contained within the 95% confidence bounds. Therefore, we cannot reject the joint hypothesis (A2, A3).

Finally, we analyse the non-take up of the reform (noncompliance) along the lines of section 5.2. To do so, we compare the IV estimate  $\hat{TE}(6, a)$  with an estimate of the *NTE* (3.5), see the dashed line in figure 3a. At all points at which the *TE* is positive, the estimated *NTE* is smaller than the corresponding estimated *TE*. The difference is significant at the 95% level.

Based on these findings, the following conclusions can be drawn. First, the policy reform increased the transition rate out of unemployment. Economically, our finding contributes to a better understanding of the relative response of individuals to monetary vs. non-monetary incentives. Second, ignoring endogeneity that arises from noncompliance leads to a negative bias in the estimates. The main implication here is that there are many good risks among the noncompliers. Thus, it is plausible to conclude that the non-take up of the reform is driven by individuals who expect to soon find a job. This finding supports the non-take up analysis by [Blasco \(2009\)](#).

APPENDIX A. IDENTIFICATION AND ESTIMATION OF ADDITIVE TREATMENT EFFECTS ON THE  
HAZARD

Denote with  $\theta_{T(s)}(t | X, V)$  the hazard of  $T(s)$  at  $t$  for an individual with characteristics  $X$  and  $V$ ,  $\theta(t | X, V) := \lim_{dt \rightarrow 0} P(T \in [t, t + dt | T \geq t, X, V])/dt$  (all expressions are assumed to exist). Then the individual additive treatment effect on the hazard at  $t$  is defined as

$$(A.1) \quad \theta_{T(t)}(t | X, V) - \theta_{T(t')}(t | X, V),$$

where  $t' > t$ . Similarly to the case with the conditional survival function, the average treatment effect on the hazard (HTE) is defined as

$$(A.2) \quad HTE(t, t') := \mathbb{E}[\theta_{T(t)}(t | S(t) = t, X, V) - \theta_{T(t')}(t | S(t) = t, X, V) | T(t) \geq t, S(t) = t, X].$$

In this subsection, we state conditions under which HTE is identified. We also develop the estimation theory. The HTE deserves a special attention for two reasons. First, the hazard of the duration variable represents the most interesting feature of its distribution in multiple applications, see [Van den Berg \(2001\)](#) for various examples and a discussion. Second, estimation of hazard effects in a treatment evaluation framework involves estimation at the boundary of the admissible domain. We develop an estimator that takes into account the region of estimation and does not lead to an increased bias.

**A.1. Identification.** Write  $W = (X, V)$  and let  $\Omega_W$  be the set of possible values for  $W$ . Further, write  $\Psi(t | X) := HTE(t, X)$ . Again we assume access to an i.i.d. sample

$$(\tilde{T}_1, S_1, Z_1, X_1, \delta_1), \dots, (\tilde{T}_n, S_n, Z_n, X_n, \delta_n).$$

The following mild technical assumption ensures that the order of taking the limit and the expectation operator can be interchanged.

**Assumption HTE1.** There exists a measurable function  $g : \mathbb{R}^+ \times \Omega_W \rightarrow \mathbb{R}^+$  that fulfills  $\mathbb{E}[g(t, W)] < \infty$  and  $|\theta(t | W = w)| \leq g(t, w)$  for each  $(t, w) \in \mathbb{R}^+ \times \Omega_W$ .

Identification is stated in the following result.

**Proposition A.1.** *Under assumptions A1-A5 and HTE1,  $\Psi(t | X)$  is identified and it holds*

$$(A.3) \quad \Psi(t | X) := \frac{\theta(t | X, Z = t) - \theta(t | X, Z = t')}{P(S = t | T \geq t, X, Z = t)}.$$

*HTE is the limit case of the general treatment effect TE,  $HTE = \lim_{n \rightarrow 0} TE/dt$ .*

**A.2. Estimation.** Henceforth, we denote with  $\theta_1(t | X)$  the hazard  $\theta(t | X, Z = t)$  of the younger cohort,  $\{Z = t\}$ , and with  $\theta_2(t | X)$  the hazard  $\theta(t | X, Z = t')$  of the older cohort. If the treatment is effective, then there will be a jump in the hazard function at the moment of treatment (per definition). Hence, when estimating  $\Psi(t | X)$ , only the observations  $\tilde{T}$  that are bigger than or equal to  $t$  are informative about  $\theta_1(t | X)$ .<sup>23</sup> This leads to estimating a hazard at the left boundary of the interval  $[t, \bar{T})$  where  $\bar{T}$  is some maximum duration, possibly  $\infty$ . Smooth hazard estimators that use a symmetric kernel would have a large bias at  $t$ , a problem called boundary effect in the literature, [Müller and Wang \(1994\)](#). Without loss of generality, let  $[0, 1]$  be the set of possible values of the duration variable and  $b = b(n)$  a bandwidth of a kernel estimator,  $b < 0.5$ . The set  $B_L := \{t : 0 \leq t < b\}$  is called a left boundary region (we do not discuss problems arising at the right boundary here). Employing a symmetric kernel to estimate the hazard at a point from that region could lead to a high bias, because the support of the kernel exceeds the range of the data. In the interior  $(0, 1)$ , this is only a finite sample problem. At the boundary  $t = 0$ , the problem persists with increasing sample size  $n$ . Boundary problems are not endemic to hazards, they arise also in the estimation of a density function, see [Karunamuni and Alberts \(2005\)](#). [Müller and Wang \(1994\)](#) develop a class of asymmetric kernels and use them to adapt the

---

<sup>23</sup>This does not apply to  $\theta_2(t | X)$ .

unconditional Ramlau-Hansen estimator to the boundary case. The kernels vary with the point of estimation and have a support that does not exceed the range of the duration variable. These kernels are referred to as boundary kernels. Following this approach, we adapt the conditional kernel hazard estimator of [Nielsen and Linton \(1995\)](#) to the case of estimation at the boundary by using boundary kernels. For simplicity, we assume that we estimate  $\Psi(t | x)$  at an interior point  $x$  of  $\Omega_X$ . Let  $k$  be a symmetric one-dimensional continuous density function with support  $[-1, 1]$ , that is

$$\int_{-1}^1 k(y)dy = 1 \quad \text{and} \quad \int_{-1}^1 yk(y)dy = 0$$

and define  $k_1$  and  $k_2$  as

$$k_1 = \int_{-1}^1 y^2 k(y)dy \quad \text{and} \quad k_2 = \int_{-1}^1 k^2(y)dy.$$

Define the  $q$ -dimensional product kernel  $K(x) = \prod_{i=1}^q k(x(i))$ , where  $x = (x(1), \dots, x(q))$ .

Next, let  $k_+$  denote the asymmetric kernel function

$$k_+ : [0, 1] \times [-1, 1] \rightarrow \mathbb{R}$$

$$(h, y) \rightarrow \frac{12}{(1+h)^4} (y+1)[y(1-2h) + (3h^2 - 2h + 1)/2].$$

This is a boundary kernel function as defined in [Müller and Wang \(1994\)](#).<sup>24</sup> The support of  $k_+(h, \cdot)$  is  $[-1, h]$ . In analogy to the symmetric kernel  $k$ , we define the second moments of  $k_+(0, \cdot)$  as

$$k_1^+ = \int_{-1}^0 y^2 k_+(0, y)dy \quad \text{and} \quad k_2^+ = \int_{-1}^0 k_+^2(0, y)dy.$$

Using standard counting processes notation, define for  $i = 1, \dots, n$  the observed failure process of the  $i^{\text{th}}$  individual at time  $t$ ,  $N_i(t) := 1\{\tilde{T}_i \leq t, T_i \leq C_i\}$  and the individual process at risk,  $Y_i(t) := 1\{\tilde{T}_i \geq t\}$ . To differentiate between observations from the cohorts 1, that is  $\{Z = t\}$ , and 2, that is  $\{Z = t'\}$ , we add a subscript 1 or 2, respectively. For example,

---

<sup>24</sup> An alternative approach could be to use the boundary kernels by [Cattaneo et al. \(2017\)](#).

$X_{1,i}$  denotes an observation of  $X$  that comes from the cohort  $\{Z = t\}$ . Then our estimator  $\widehat{\Psi}(t | x)$  of  $\Psi(t | x)$  is defined as

$$(A.4) \quad \widehat{\Psi}(t | x) := \frac{1}{\hat{p}_1(t | x)} \left( \frac{\sum_{i=1}^n K\left(\frac{x-X_{1,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) dN_{1,i}(s)}{\sum_{i=1}^n K\left(\frac{x-X_{1,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) Y_{1,i}(s) ds} - \frac{\sum_{i=1}^n K\left(\frac{x-X_{2,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) dN_{2,i}(s)}{\sum_{i=1}^n K\left(\frac{x-X_{2,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) Y_{2,i}(s) ds} \right),$$

where  $\hat{p}_1(t | x)$  is a nonparametric estimator for  $p_1(t | x) := P(S = t | T \geq t, X = x, Z = t)$ . We assume that  $\hat{p}_1(t | x)$  is consistent. In addition, for proposition A.2 ii) we assume that  $b^{-2}(\hat{p}_1(t | x) - p_1(t | x)) = o_p(1)$ , which can be assured by assuming that  $p_1(t | x)$  is sufficiently smooth in  $x$ . The term

$$\hat{\theta}_j(t | x) := \frac{\sum_{i=1}^n K\left(\frac{x-X_{j,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) dN_{j,i}(s)}{\sum_{i=1}^n K\left(\frac{x-X_{j,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) Y_{j,i}(s) ds}$$

for  $j = 1, 2$  is a conditional smooth hazard estimator for  $\theta_j(t | x)$  developed in [Nielsen and Linton \(1995\)](#) and adapted to the boundary case. Define

$$(A.5) \quad \theta_j^*(t | x) := \frac{\sum_{i=1}^n K\left(\frac{x-X_{j,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) \theta_j(s | X_{j,i}) Y_{j,i}(s) ds}{\sum_{i=1}^n K\left(\frac{x-X_{j,i}}{b}\right) \int k_+(t, \frac{t-s}{b}) Y_{j,i}(s) ds} \quad j = 1, 2$$

and

$$(A.6) \quad \Psi^*(t | x) = \frac{1}{\hat{p}_1(t | x)} (\theta_1^*(t | x) - \theta_2^*(t | x)).$$

We need the following assumptions.

H1  $\mathbb{E}[Y_i(s)] = u(s)$  and  $u(\cdot)$  is continuous

H2 i)  $f(x)u(t)$  is positive on a neighborhood  $U$  of  $(0, x_0) \in \mathbb{R}^+ \times \Omega_X$ , where  $x_0$  is an interior point of  $\Omega_X$  and  $f$  is the density of  $X$ . ii)  $\theta_j$  is twice continuously differentiable on  $U$ . iii)  $fu$  is continuously differentiable on  $U$ .

H3  $nb^{q+1} \rightarrow \infty$  and  $b = b(n) \rightarrow 0$  as  $n \rightarrow \infty$ .

The following proposition states the pointwise asymptotic properties of  $\widehat{\Psi}(0 | x_0)$ .

**Proposition A.2.** *Define*

$$\sigma_{\Psi}^2 := k_2^+ k_2^q \frac{1}{p_1^2(0 | x_0)} (\theta_1(0 | x_0)/f_1(x_0) + \theta_2(0 | x_0)/f_2(x_0)).$$

*Under assumptions H1-H3, the following results hold:*

i)  $\sqrt{nb^{q+1}}(\widehat{\Psi}(0 | x_0) - \Psi^*(0 | x_0)) \xrightarrow{d} N[0, \sigma_{\Psi}^2].$

ii) *If in addition  $b^{-2}(\hat{p}_1(t | x) - p_1(t | x)) = o_p(1)$ , then*

$$\begin{aligned} b^{-2}(\Psi^*(0 | x_0) - \Psi(0 | x_0)) &\xrightarrow{p} \sum_{i=1}^2 \frac{(-1)^{i+1} k_1^+}{f_i(x_0) u_i(0) p_1(0 | x_0)} \\ &\left[ \frac{\partial \theta_i(0 | x_0)}{\partial t} \frac{\partial (f_i(x_0) u_i(0))}{\partial t} + \frac{1}{2} \frac{\partial^2 \theta_i(0 | x_0)}{\partial t^2} f_i(x_0) u_i(0) + \right. \\ &\left. \sum_{j=1}^q \left( \frac{\partial \theta_i(0 | x_0)}{\partial x(j)} \frac{\partial (f_i(x_0) u_i(0))}{\partial x(j)} + \frac{1}{2} \frac{\partial^2 \theta_i(0 | x_0)}{\partial x(j)^2} f_i(x_0) u_i(0) \right) \right] \end{aligned}$$

iii) *Finally, it also holds*

$$\begin{aligned} \hat{\sigma}_{\Psi}^2 &:= \frac{nb^{q+1}}{\hat{p}_1(0 | x_0)} \sum_{j=1}^2 \frac{\sum_{i=1}^n K^2\left(\frac{x_0 - X_{ji}}{b}\right) \int k_+^2\left(\frac{-s}{b}\right) dN_{ji}(s)}{\left(\sum_{i=1}^n K\left(\frac{x_0 - X_{ji}}{b}\right) \int k_+\left(\frac{-s}{b}\right) Y_{ji}(s) ds\right)^2} \xrightarrow{p} \\ &\sigma_{\Psi}^2 \end{aligned}$$

Result i) gives the asymptotic distribution of the estimator, ii) characterizes the bias and iii) provides the standard errors for confidence bounds around  $\Psi^*$ . If the bandwidth is chosen to be of  $o(n^{-1/(q+5)})$ , then the asymptotic bias is negligible and proposition A.2 can be used to construct confidence bands for  $\Psi$ .

## APPENDIX B. EMPIRICAL APPLICATION: DATA DESCRIPTION AND ADDITIONAL RESULTS

**B.1. Dataset and empirical strategy.** The dataset we use is constructed by matching two administrative data sets: the Fichier Historique (FH) dataset, which contains information about the unemployment spells and is issued by the French public employment agency (Agence Nationale Pour L'emploi, ANPE), and the Déclaration Anuelle de Données Sociales (DADS) dataset, which contains the employment information of all individuals employed in the private sector and is issued by the French Statistical Institute (INSEE).

We extract a set of variables, rich enough to account for the socioeconomic status of the individuals, namely age, gender, marital status, number of children, educational level, professional experience, reason for entering unemployment, exit direction (out of unemployment), and unemployment history. Details about the construction and content of the variables are provided below in section B.4.

To preclude geographical heterogeneity we restrict our sample to the administrative region Île de France, which contains Paris and consists of the administrative departments 75, 77, 78, 91, 92, 93, 94 and 95. Because of its size and specific infrastructure, this region might differ from the rest of France in terms of labor market dynamics (mobility, unemployment structure, wages) and in terms of the implementation of the reform. Moreover, the macroeconomic conditions in this region are stable over the period of consideration, which ensures the comparability of the cohorts, see subsection B.3.1.

The choice of the cohorts is restricted by the available data. There is no administrative variable that captures the compliance status of the unemployed. We develop a novel approach to deal with this problem, which so far has not been adopted in other PARE evaluation studies with register data. Specifically, we choose the younger (i.e., treated) cohort  $\{Z = t\}$  such that its first due benefits reduction under the old system coincides with the implementation of the reform. This enables us to observe the compliance status.<sup>25</sup> Its inflow is six months before the start of PARE.<sup>26</sup> The choice of the comparison cohort (the untreated) is more flexible as we do not need to observe the compliance. The main

---

<sup>25</sup>One may also consider subsequent elapsed durations at which declines take place, but this would be at the cost of having fewer observations.

<sup>26</sup>The time length from inflow until the day of the first decline can vary somewhat, depending on characteristics of the unemployed, such as number of working days in the last twelve months and age; see [Freyssinet \(2002\)](#) for details. We stop the duration clock on days on which the individual worked part-time during their unemployment spell. Excluding them does not affect the results. We also exclude elderly unemployed.

concern is to prevent cohort effects. Business cycles or mass layoffs due to bankruptcies of large firms are examples for possible causes for structural changes in the inflow over time. We choose the comparison cohort to have entered unemployment 3 months earlier than the treated cohort because then both cohorts begin their unemployment spells in a fairly economically stable time interval; see subsection B.3.1 for a discussion. This choice has an implication for the time interval of comparison. Conditional on survival up to 6 months, one can compare the two cohorts only in an interval of 3 months. After the 3rd month, the older cohort will also receive the treatment, and one would no longer compare treated with untreated.

With these choices we end up with 537 (311) spells in the treated (comparison) cohort. From these, 116 (76) are censored. In the treated cohort there are 250 compliers.

**B.2. Treatment effect heterogeneity.** In this section, we present estimation results on the treatment effect for two different subgroups. In particular, figure 4a displays the estimates for white versus blue collars, while figure 4b displays the results for higher (above high school) and lower educated. Both figures reveal patterns that are very similar to the

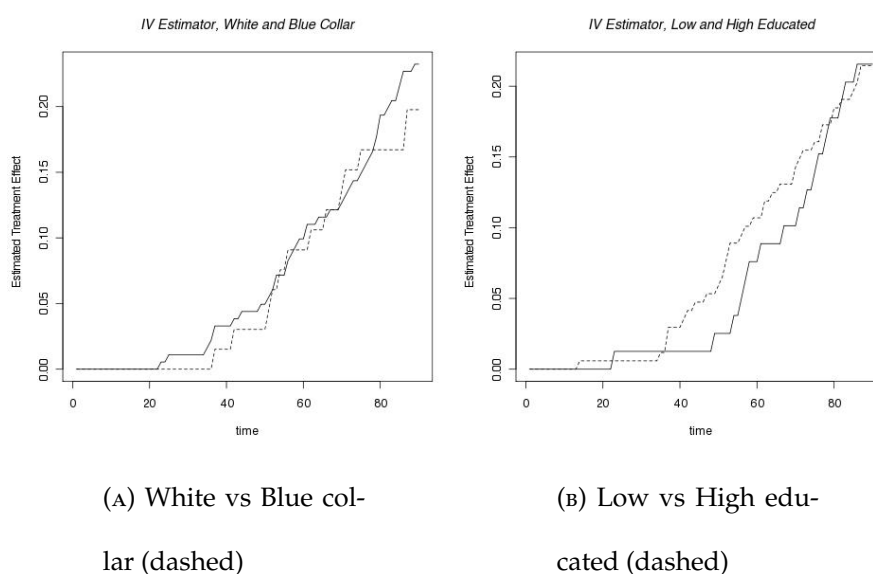


FIGURE 4. Estimates conditional on qualification and education



unconditional estimates in the paper.

### B.3. Model diagnostics.

B.3.1. *Cohort effects.* In this subsection, we study the comparability of the cohorts. Cohort effects would violate the randomization assumption A3. Since  $T(s)$ ,  $S(t)$  and  $V$  are unobserved, A3 cannot be tested directly. An indirect way to assess the plausibility of A3, in addition to the joint test of A2 and A3 in the last section of the main paper, is (i) to test whether pre-treatment characteristics are balanced at inflow and (ii) to assess the macroeconomic conditions at the point in time of inflow of the two cohorts.

We first perform a chi-square test for equality of distributions of level of education, years of experience, number of children, gender and pre-jobloss wages. The corresponding p-values are 0.6037, 0.98, 0.5112, 0.581, and 0.34, which indicates that the differences between these distributions are statistically insignificant. Second, the same test is performed also for the layoff reasons. The null (equality of distributions) is rejected, but in this case this could be due to the large number of categories and small number of observations in each category. A histogram of aggregated categories indicates that the cohorts are indeed similar, see figure 5a-5b. Third, the average level of unemployment in the administrative region Îll de France in the first three quarters of 2001 is constant and equal to 6.4%, which is evidence for a fairly stable macroeconomic environment.<sup>27</sup>

Finally, we challenge the "no cohort effects" assumption hidden in A3 with an alternative choice of a control cohort. Figure 6 shows a comparison of the estimates of the treatment effect with two different control cohorts. The thick line is the estimate with the 9-months old cohort (3 months older than the treatment cohort), while the dashed line displays an estimate with an 8-month old cohort (i.e. 2 months older than the treatment cohort). The estimates are very similar. The main conclusion from this analysis is that the

---

<sup>27</sup> Source: <http://www.insee.fr/en/bases-de-donnees/bsweb>

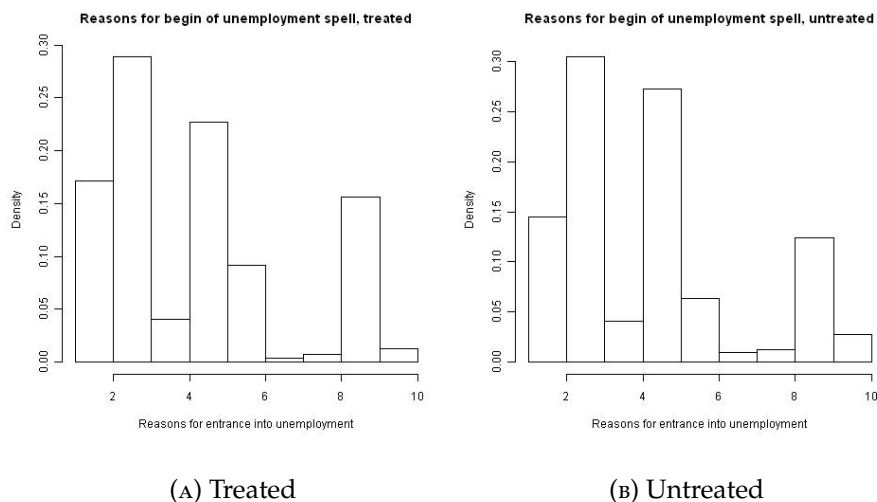


FIGURE 5. Histograms of layoff reasons

results are not sensitive with respect to the choice of the control, which is further evidence for the plausibility of our assumptions.<sup>28</sup>

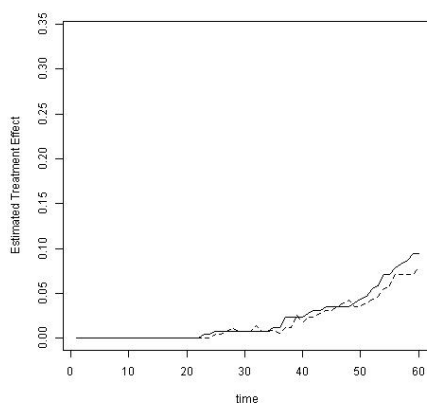


FIGURE 6. IV estimates with (i) 9-month-old cohort (thick line) and (ii) 8-month-old-cohort (dashed line)

B.3.2. *No anticipation.* Next, the “no anticipation” assumption is fulfilled when individuals do not anticipate the moment in time of treatment or do not act upon this information, see for a discussion [Abbring and van den Berg \(2003\)](#). Although it was known that a reform is going to take place, there was a lot of debate and uncertainty over its content.

<sup>28</sup> We are thankful to an anonymous referee for suggesting this sensitivity analysis to us.

Unemployed were informed about the exact content and launch date on the 18th of June 2001, that is, less than two weeks before the start of the program, so they had practically no time to react upon this information, see [Freyssinet \(2002\)](#). Further, when an individual decides to switch to the new system, the assignment to a specific treatment depends mostly on the social worker in charge and on the slots available, so that the unemployed has no knowledge of it in advance, see also [Crépon et al. \(2005\)](#). Combined with a very short time span between assignment and launch of a treatment is very short, which precludes acting upon the anticipation. In the main paper, we test for equality of pre-treatment outcomes, which is an implication of (jointly) A2 and A3.

*B.3.3. Dependent Censoring: a Simulation Study.* The last important assumption is that of independent censoring. It cannot be tested directly, as revealed by a nonidentification result of [Tsiatis \(1975\)](#). Over 70% of all censored spells are attributed to the censoring categories “no control”, “other cases” and “other termination of search”. There is no further information for these cases.

We therefore assess the impact of the assumption of independent censoring in an indirect way: we conduct a small simulation study. Deviations from  $C \perp S$  and  $C \perp T$  are constructed, where  $C$  again is a censoring random variable. The first one influences the estimator of the probability to be a complier,

$$P(S = t \mid T \geq t, X, Z = t),$$

while the second one influences the estimator of the difference

$$P(T \in [t, t + a) \mid T \geq t, X, Z = t) - P(T \in [t, t + a) \mid T \geq t, X, Z = t').$$

We are interested in their marginal impacts as well as in the influence of their interplay. Two cohorts are simulated, the treated and the nontreated, each with 10000 individuals. Both cohorts consist of compliers and noncompliers and in each cohort the probability to

be a complier is 80%. Noncompliers dominate stochastically the compliers when both groups have not received the treatment. This reflects our finding in section 5.2 that noncompliance might occur due to the expectation of a short spell. The treatment is obtained by the compliers of the first cohort on the 20th day after inflow and it shifts their duration distribution from  $N(60, 15)$  to  $N(30, 10)$  in line with the estimation results from section 6<sup>29</sup>. The noncompliers are not influenced by the treatment and have a duration distribution  $N(45, 15)$ . The compliers from the second cohort do not receive the treatment too. Their duration distribution is equal to the duration distribution of the compliers of cohort 1 before treatment,  $N(60, 15)$ . Figure 7 shows the theoretical treatment effect,

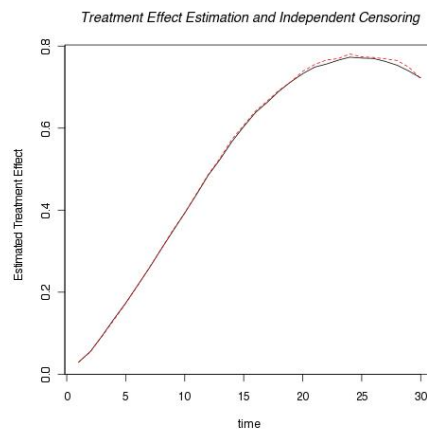


FIGURE 7. An IV estimator of the treatment effect. Time measured in days.

Day 0 corresponds to the day of treatment (day 20).

depicted by the thick black line. The dashed red line represents the IV estimator in a case with independent censoring with a distribution  $N(40, 10)$  (the second argument is henceforth the standard deviation) . This is the benchmark estimator.

Next, a dependence of the censoring on the compliance is introduced. The different choices of distributions are described in table 1.

<sup>29</sup>Negative values are replaced by their absolute values.

TABLE 1. Simulation of dependences between censoring and compliance

Line description	Censoring distribution compliers	Censoring distribution noncompliers
Green dashed line	N(30,15)	N(50,15)
Red dotted line	N(30,15)	N(40,15)
Blue long dashed line	N(40,15)	N(30,15)
Grey two dashed line	N(50,15)	N(30,15)

Notes: The second argument of the normal distribution is its standard deviation

The resulting estimators are shown in figure 8. The solid black line is theoretical effect.

The figure reveals the relationship between bias of the treatment effect and dependence

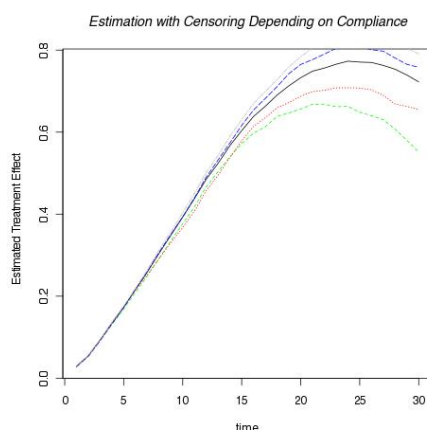


FIGURE 8. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical treatment effect. Different curves correspond to different dependences of censoring and compliance, see table 1. The solid black line is theoretical effect.

of censoring and compliance. When the compliers are at higher risk of censoring, the treatment effect is (a. e.) underestimated. The higher this discrepancy in the risk exposure,

the bigger the bias. Similarly, when the noncompliers are at higher risk of censoring, the treatment effect is overestimated.

Next, the relationship between bias and time dependence of the censoring is exploited. We simulate three different levels of dependence. In all three cases long spells have a higher risk of being censored than short spells. This is in line with typical situations in applied survival analysis. For example, long term unemployed might have smaller incentives to meet criteria (e. g. administrative control of search, regular visits at the agency, etc.) to stay on an unemployment insurance list. The three specifications are defined in table 2. Each row represents one specification.

TABLE 2. Simulation of dependences between censoring and time

Line description	Censoring distribu- tion $T \leq 40$	Censoring distribu- tion $T > 40$
Green dashed line	N(40,20)	N(30,20)
Red dotted line	N(40,20)	N(25,20)
Blue long dashed line	N(40,20)	N(20,20)

Notes: The second argument of the normal distribution is its standard deviation

The corresponding estimators are depicted in figure 9. Approximately until day 15 the IV estimator performs fairly well in all three cases. Afterwards it underestimates the treatment effect. The bias increases in absolute value with increasing time dependence (defined as the difference in the means in the two groups of spells).

It is interesting to simulate and analyze a combination of these two types dependence patterns. We simulate four patterns of such an interplay. The concrete distributions are described in table 3. The results are shown in figure 10. The blue and the grey lines are closer to the theoretical effect than the other two estimators. This indicates, that a violation

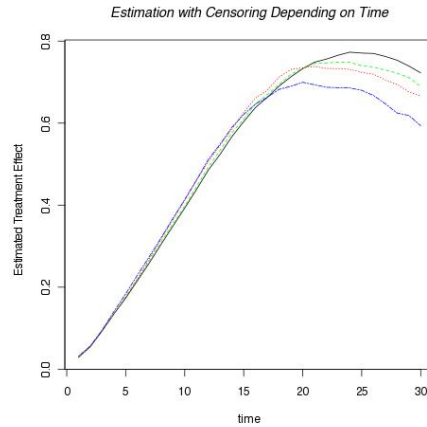


FIGURE 9. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical treatment effect. Different curves correspond to different dependence patterns of censoring and time, see table 2. The solid black line is theoretical effect.

TABLE 3. Simulation of dependences between censoring and compliance and time

Line description	$K, T \leq 30$	$K, T > 30$	$N, T \leq 30$	$N, T > 30$
Green dashed line	N(50,20)	N(30,20)	N(30,20)	N(20,20)
Red dotted line	N(40,20)	N(30,20)	N(30,20)	N(20,20)
Blue two dashed line	N(30,20)	N(20,20)	N(40,20)	N(30,20)
Grey long dashed line	N(30,20)	N(20,20)	N(50,20)	N(30,20)

Notes: K stays for compliers, N for noncompliers.

in the censoring assumption  $C \perp S$  might partially offset a violation in the assumption  $C \perp T$ . This is a novel result.

In the French labor market reform it is difficult to argue which type of dependence there is likely to be. Noncompliers contain many quick exits, and if longer spells have a higher censoring risk than shorter spells, than noncompliers should be less exposed to

censoring than compliers. This would correspond to the fourth case of table 3. Thus the simulation results provide evidence, that the IV estimator is robust to a violation in the independent censoring assumption.

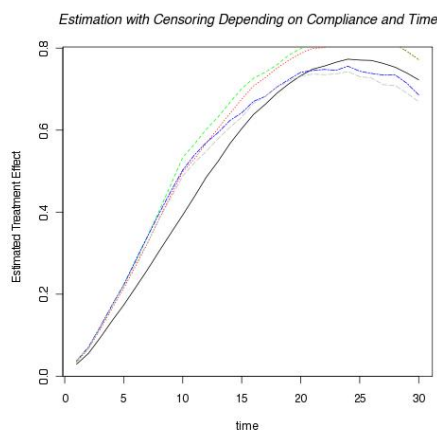


FIGURE 10. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical effect in the absence of censoring. Different curves correspond to different dependences of censoring and time, see table 2. The solid black line is theoretical effect.

**B.4. Description of variables.** The variables used in our empirical application have been constructed in the following way:

- The variable **age** gives the age at the begin of the unemployment spell and is defined as the year in which the spells begins minus the year of birth.
- **Marital status** consists of four categories: single, married, divorced and widowed.
- the variable for **educational level** summarizes the 31 categories used in the administrative data set into 6 categories according to the highest degree attained. The correspondence is roughly as follows: value 1 if the degree is in niveau I and II (university degree, maîtrise and licence), value 2 if the degree is in niveau III - BTS and DUT (brevet de technicien supérieur and diplôme universitaire de technologie,



respectively, both technical degrees obtained in 2 years after high school), value 3 for all Baccalauréat (high school degree, the general part of lycée) diplomas and for all dropouts from niveau III, 4 for all BEP ,CEP (professional Baccalauréat, specialised part of lycée) and all dropouts from Baccalauréat, 5 for BEPC (brevet d'études du premier cycle, junior high school), and 6 for below.

- The variable **experience** states the number of years of experience in the job (type and position), which the individual is looking for. The types of jobs are specified in an administrative nomenclature table (ROME table). There are several hundred different types.
- The **job type** variable contains general information about the type of the activity in the job preceding the current unemployment spell. It summarizes the 9 administrative categories into 6 categories: white collar skilled, white collar unskilled, technical, supervisor (a production team leader) and manager. This summarized categorization is in line with existing literature, see for example ?. The initial administrative variable is contained in the FH data set. This holds also for the variable, which states which job is the unemployed looking for, while the following employment type and position is contained in the DADS data set. Unfortunately, there is no clear matching between the variables from the two different data sets, which leads to some unclarity regarding the question whether the unemployed actually found the job he/she was looking for. This restricts our definition of censoring. Therefore, in this application each observation with known job destination is considered uncensored.
- **Censoring indicator:** there are several possibilities, when an observation is considered as censored. These are:
  - when the unemployment spell in the data set is not finished at the time of the data collection, or

- when the individual exits the labor market. This includes exits to maternity, accident, illness or invalidity, invalidity pension, military service, administrative change of insurance status, attrition because of insufficient administrative control, dropout because of irregular notifications, and other, unspecified reasons. While reasons such as maternity, military services and invalidity pension are normally known well in advance by the unemployed and can therefore be related to search activity (as well as to compliance behavior), they represent a small fraction of the observations.
- **Unemployment history:** it is constructed as a binary variable which equals 1 if the individual had been already unemployed before the last employment spell. There are various ways to define unemployment history. One example is the total length of previous unemployment spells. Alternatively, one could take the number of unemployment spells, or both. All possibilities suffer from disadvantages. The last possibility seems to provide the most complete information, but it also demands more data, since it provides many different categories. The total length of previous unemployment lacks any information about the lengths of the separate spells, and the number of spells alone doesn't give any information about the length of unemployment. The binary indicator also does not provide any information at all about the dispersion of previous unemployment, but it is easy to understand and requires only two categories, which makes it computationally attractive. Additional, more serious drawback for the other two indicators is, that the data set is left censored: the earliest information about employment is from 1993. This problem is less severe, if one only looks at the indicator of having been unemployed.

## APPENDIX C. PROOFS OF PROPOSITIONS

## C.1. Proofs of propositions in section 3.2.

*Proof of proposition 3.1.* First we show that from the no anticipation assumption the following result holds:

$$(C.1) \quad P(T(t) \geq t \mid X, S(t) = t) = P(T(t') \geq t \mid X, S(t) = t).$$

This is so because

$$\begin{aligned} P(T(t) \geq t \mid X, S(t) = t, V) &= \exp(-\Theta_{T(t)}(t \mid X, S(t) = t, V)) \\ &\stackrel{\text{No anticipation}}{=} \exp(-\Theta_{T(t')}(t \mid X, S(t) = t, V)) = P(T(t') \geq t \mid X, S(t) = t, V), \end{aligned}$$

so that we obtain

$$\begin{aligned} P(T(t) \geq t \mid X, S(t) = t) &= \mathbb{E} [I_{\{T(t) \geq t\}} \mid X, S(t) = t] \\ &= \mathbb{E} [\mathbb{E} [I_{\{T(t) \geq t\}} \mid X, S(t) = t, V] \mid X, S(t) = t] \\ &= \mathbb{E} [P(T(t) \geq t \mid X, S(t) = t, V) \mid X, S(t) = t] \\ &= \mathbb{E} [P(T(t') \geq t \mid X, S(t) = t, V) \mid X, S(t) = t] \\ &= \mathbb{E} [\mathbb{E} [I_{\{T(t') \geq t\}} \mid X, S(t) = t, V] \mid X, S(t) = t] = P(T(t') \geq t \mid X, S(t) = t) \end{aligned}$$

where  $I_{\{T(s) \in B\}}$  is an indicator function equal to 1 when  $T(s) \in B$  (of course from these steps we also see that  $P(T(t) \geq t \mid X, S(t) = t, V) = P(T(t') \geq t \mid X, S(t) = t, V)$ ).

Next, using result (C.1), we show  $F_{V|T(t) \geq t, X, S(t)=t} = F_{V|T(t') \geq t, X, S(t)=t}$ . Let  $B$  be a Borel set.

With result (C.1), it holds

$$P(V \in B \mid T(t') \geq t, X, S(t) = t) = P(V \in B \mid T(t) \geq t, X, S(t) = t).$$

Now we show  $F_{V|T(t) \geq t, X, S(t)=t} = F_{V|T \geq t, X, S=t, Z=t}$ . First we observe that  $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X, V$  and  $Z \perp\!\!\!\perp V \mid X$  together imply  $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X$  (Weak Union, see [Pearl \(2000\)](#)). Then,

we have

$$P(V \in B \mid T(t) \geq t, X, S(t) = t) = \frac{P(V \in B \mid X, S(t) = t)P(T(t) \geq t \mid X, S(t) = t, V \in B)}{P(T(t) \geq t \mid X, S(t) = t)}.$$

We study the separate components of the right-hand side of the last expression.

(1) With assumptions A3 and A4, it holds

$$P(V \in B \mid X, S(t) = t) = P(V \in B \mid X, S = t, Z = t).$$

(2) Further,

$$P(T(t) \geq t \mid X, S(t) = t, V \in B) = P(T \geq t \mid X, S = t, V \in B, Z = t).$$

(3) Using  $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X$  instead of  $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X, V$ , we obtain

$$P(T(t) \geq t \mid X, S(t) = t) = P(T \geq t \mid X, S = t, Z = t)$$

So finally we get the equality

$$\begin{aligned} & P(V \in B \mid T(t) \geq t, X, S(t) = t) \\ &= \frac{P(V \in B \mid X, S = t, Z = t)P(T \geq t \mid X, S = t, V \in B, Z = t)}{P(T \geq t \mid X, S = t, Z = t)} \\ &= P(V \in B \mid T \geq t, X, S = t, Z = t) \end{aligned}$$

□

*Proof of corollary 3.1.* With proposition 3.1,

$$\begin{aligned} TE(t, t', a) &= \mathbb{E}\left[P(T(t) \in [t, t+a) \mid T(t) \geq t, X, V, S(t) = t) \mid T(t) \geq t, X, S(t) = t\right] \\ &- \mathbb{E}\left[P(T(t') \in [t, t+a) \mid T(t') \geq t, X, V, S(t) = t) \mid T(t') \geq t, X, S(t) = t\right] \\ &= P(T(t) \in [t, t+a) \mid T(t) \geq t, X, S(t) = t) - P(T(t') \in [t, t+a) \mid T(t') \geq t, X, S(t) = t). \end{aligned}$$

□

**Lemma C.1.** Set  $B = [t, t + a)$  where  $a \leq t' - t$ . Under Assumptions A1-A4, it holds for all  $\infty \geq t' \geq t \geq 0$  that

(C.2)

$$P(T(t) \in B \mid T(t) \geq t, X, S(t) = t) = P(T \in B \mid T \geq t, X, S = t, Z = t),$$

(C.3)

$$P(T(t') \in B \mid T(t') \geq t, X, S(t) = \infty) = P(T \in B \mid T \geq t, X, S = \infty, Z = t) \quad \text{and}$$

(C.4)

$$P(T(t') \in B \mid T(t') \geq t, X) = P(T \in B \mid T \geq t, X, Z = t').$$

*Proof of Lemma C.1.* First, observe that with randomization and consistency, it holds

$$P(T(t) \in B \mid X, S(t) = t) = P(T \in B \mid X, S = t, Z = t),$$

$$P(T(t) \geq t \mid X, S(t) = t) = P(T \geq t \mid X, S = t, Z = t),$$

so that

$$P(T(t) \in B \mid T(t) \geq t, X, S(t) = t) = P(T \in B \mid T \geq t, X, S = t, Z = t)$$

where the r.h.s of the equality consists only of observables.

Next, we have

$$\begin{aligned} & P(T \in B \mid X, S = \infty, Z = t) = P(T(\infty) \in B \mid X, S = \infty, Z = t) \\ &= P(T(\infty) \in B \mid X, S(t) = \infty, Z = t) = P(T(\infty) \in B \mid X, S(t) = \infty) \\ &= P(T(t') \in B \mid X, S(t) = \infty), \end{aligned}$$

where the first and the second equalities follow due to consistency, the third due to randomization and the fourth due to no anticipation. Equality (C.4) follows analogically.

□

**Lemma C.2.** *Under Assumptions A1-A4, it holds for all  $\infty \geq t' \geq t \geq 0$  that*

$$(C.5) \quad P(S(t) = t \mid T(t) \geq t, X) = P(S = t \mid T \geq t, X, Z = t),$$

$$(C.6) \quad P(S(t) = t \mid T(t') \geq t, X) = P(S(t) = t \mid T(t) \geq t, X).$$

*Proof of Lemma C.2.* First, it holds

$$\begin{aligned} P(S = t \mid T \geq t, X, Z = t) &= \frac{P(T \geq t \mid S = t, X, Z = t)P(S = t \mid X, Z = t)}{P(T \geq t \mid X, Z = t)} \\ &= \frac{P(T(t) \geq t \mid S(t) = t, X)P(S(t) = t \mid X)}{P(T(t) \geq t \mid X)} = P(S(t) = t \mid T(t) \geq t, X), \end{aligned}$$

where the second equality follows with assumptions A1-A4.

Next,

$$\begin{aligned} P(S(t) = t \mid T(t') \geq t, X) &= \frac{P(S(t) = t, T(t') \geq t \mid X)}{P(T(t') \geq t \mid X)} \\ &= \frac{P(T(t') \geq t \mid S(t) = t, X)P(S(t) = t \mid X)}{P(T(t) \geq t \mid X)} = \frac{P(T(t) \geq t \mid S(t) = t, X)P(S(t) = t \mid X)}{P(T(t) \geq t \mid X)} \\ &= P(S(t) = t \mid T(t) \geq t, X), \end{aligned}$$

where the second equality holds due to no anticipation. □

*Proof of proposition 3.2.* First, write

$$\begin{aligned} (C.7) \quad &P(T(t') \in [t, t+a) \mid T(t') \geq t, X) \\ &= P(T(t') \in [t, t+a) \mid T(t') \geq t, X, S(t) = t)P(S(t) = t \mid T(t') \geq t, X) \\ &+ P(T(t') \in [t, t+a) \mid T(t') \geq t, X, S(t) = \infty)P(S(t) = \infty \mid T(t') \geq t, X), \end{aligned}$$

and then express  $P(T(t') \in [t, t+a) \mid T(t') \geq t, X, S(t) = t)$  in terms of the other three components of equality (C.7). Plugging in the results of lemma C.1 and lemma C.2, we

obtain for  $F_{C,0} := P(T(t') \in B \mid T(t') \geq t, X, S(t) = t)$

$$\begin{aligned} & P(T(t') \in B \mid T(t') \geq t, X, S(t) = t) \\ = & \frac{P(T \in B \mid T \geq t, X, Z = t') - P(T \in B \mid T \geq t, X, Z = t, S = \infty)P(S = \infty \mid T \geq t, X, Z = t)}{P(S = t \mid T \geq t, X, Z = t)}. \end{aligned}$$

Finally, with  $F_{C,1} := P(T(t) \in B \mid T(t) \geq t, X, S(t) = t)$ , the treatment effect is equal to  $F_{C,1} - F_{C,0}$  which after simplification is equal to

$$\frac{P(T \in B \mid T \geq t, X, Z = t) - P(T \in B \mid T \geq t, X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)}.$$

□

*Proof of proposition 3.3.* First, note that  $P(T \in [t, t+a) \mid T \geq t, X, Z = t) = 1 - \frac{P(T \geq t+a \mid X, Z = t)}{P(T \geq t \mid X, Z = t)}$ . Each of the survival functions on the r.h.s can be consistently estimated with a Kaplan-Meier estimator. Thus,  $P(T \in [t, t+a) \mid T \geq t, X, Z = t)$  is identified. Identification of

$$P(T \in [t, t+a) \mid T \geq t, X, Z = t')$$

is shown analogously. Finally, under the independent right-censoring assumption, it holds

$$(C.8) \quad P(S = t \mid T \geq t, X, Z = t, C \geq t) = P(S = t \mid \tilde{T} \geq t, X, Z = t).$$

The expression on the r.h.s contains only observables and is also identified. This completes the proof. □

*Proof of proposition 3.4.* First, note that under assumptions A1, A2', A3', A4, A5, proposition 3.1 holds locally (i.e. for  $t' \geq t$  with  $t'$  in a  $\eta$ -neighborhood of  $t$ ). Then, locally, the treatment effect on the hazard is identified due to proposition A.1. The result of proposition 3.4 follows by taking the limit  $t' \rightarrow t^+$  of the expression  $\Psi$ . □

**C.2. Proofs of propositions in section 4.** This result follows directly from the continuity of the function  $G(a,b,c,d,e) = \frac{1}{e} \left( \frac{a}{b} - \frac{c}{d} \right)$ , the Continuous Mapping Theorem and the consistency of  $\bar{F}_i(t)$  and  $\hat{p}$ .

Define the null hypothesis

$$(C.9) \quad H_0 : \quad (\text{Ineffective treatment}) \quad \frac{\bar{F}_2(t+a)}{\bar{F}_2(t)} - \frac{\bar{F}_1(t+a)}{\bar{F}_1(t)} = 0.$$

Under (C.9), it holds

$$\begin{aligned} \sqrt{n} \widehat{TE}(t,a) &= \frac{\sqrt{n}}{\hat{p}} \left( \frac{\widehat{E}_2(t+a)}{\widehat{E}_2(t)} - \frac{\widehat{H}_1(t+a)}{\widehat{H}_1(t)} \right) = \\ &= \frac{\sqrt{n}}{\hat{p}} \left( \frac{\widehat{E}_2(t+a)}{\widehat{E}_2(t)} - \frac{\bar{F}_2(t+a)}{\bar{F}_2(t)} \right) - \frac{\sqrt{n}}{\hat{p}} \left( \frac{\widehat{H}_1(t+a)}{\widehat{H}_1(t)} - \frac{\bar{F}_1(t+a)}{\bar{F}_1(t)} \right) \end{aligned}$$

For  $i = 1, 2$  the Taylor expansion of  $\frac{\widehat{F}_i(t+a)}{\widehat{F}_i(t)}$  around  $\frac{\bar{F}_i(t+a)}{\bar{F}_i(t)}$  can be written as

$$\begin{aligned} \frac{\widehat{F}_i(t+a)}{\widehat{F}_i(t)} &= \frac{\bar{F}_i(t+a)}{\bar{F}_i(t)} + \frac{1}{\bar{F}_i(t)} (\widehat{F}_i(t+a) - \bar{F}_i(t+a)) - \frac{\bar{F}_i(t+a)}{\bar{F}_i^2(t)} (\widehat{F}_i(t) - \bar{F}_i(t)) \\ &+ O \left[ (\widehat{F}_i(t+a) - \bar{F}_i(t+a)) (\widehat{F}_i(t) - \bar{F}_i(t)) + (\widehat{F}_i(t) - \bar{F}_i(t))^2 \right], \end{aligned}$$

and therefore

$$\begin{aligned} \sqrt{n} \left( \frac{\widehat{F}_i(t+a)}{\widehat{F}_i(t)} - \frac{\bar{F}_i(t+a)}{\bar{F}_i(t)} \right) &= \frac{\sqrt{n}}{\bar{F}_i(t)} (\widehat{F}_i(t+a) - \bar{F}_i(t+a)) - \frac{\bar{F}_i(t+a) \sqrt{n}}{\bar{F}_i^2(t)} (\widehat{F}_i(t) \\ &- \bar{F}_i(t)) + O \left[ \sqrt{n} (\widehat{F}_i(t+a) - \bar{F}_i(t+a)) (\widehat{F}_i(t) - \bar{F}_i(t)) + \sqrt{n} (\widehat{F}_i(t) - \bar{F}_i(t))^2 \right]. \end{aligned}$$

The last term converges to zero in probability.

$$\text{With (4.3), the terms } \frac{\sqrt{n}}{\bar{F}_i(t)} (\widehat{F}_i(t+a) - \bar{F}_i(t+a)) \quad \text{and} \quad \frac{\bar{F}_i(t+a) \sqrt{n}}{\bar{F}_i^2(t)} (\widehat{F}_i(t) - \bar{F}_i(t))$$

are asymptotically normally distributed with mean 0 and variances

$$\frac{1}{\bar{F}_i^2(t)} \sigma_i(t+a) \quad \text{and} \quad \frac{\bar{F}_i^2(t+a)}{\bar{F}_i^4(t)} \sigma_i(t), \quad \text{respectively.}$$

The proof of the proposition follows then from the independence of the random variables

$$D_1 \text{ and } D_2, \text{ where } D_i = \frac{\widehat{F}_i(t+a)}{\widehat{F}_i(t)}, i = 1, 2.$$



### C.3. Proofs of propositions in section A of the appendix.

*Proof of proposition A.1.* Under the Lebesgue dominated convergence theorem,

$$\theta(t | X) = \lim_{dt \rightarrow 0} \mathbb{E}[P(T \in [t, t + dt) | T \geq t, X, V)/dt | T \geq t, X] = \mathbb{E}[\theta(t | X, V)],$$

and the proof follows directly from proposition 3.2.  $\square$

*Proof of proposition A.2.* For notational simplicity we drop the dependence on 0 and  $x_0$ . First note, that the results of Theorem 1 [Nielsen and Linton \(1995\)](#) remain valid at the boundary when we replace the symmetric kernel  $k$  with its boundary counterpart  $k_+$  and adapt the constants. The validity of proposition A.2 i) follows from  $\sqrt{nb^{q+1}}((\widehat{\Psi} - \Psi^*) = \frac{\sqrt{nb^{q+1}}}{\widehat{p}_1}((\widehat{\theta}_1 - \theta_1^*) - (\widehat{\theta}_2 - \theta_2^*)),$  the independence of  $(\widehat{\theta}_1 - \theta_1^*)$  and  $(\widehat{\theta}_2 - \theta_2^*),$  and the adapted proof of Theorem 1 i) in [Nielsen and Linton \(1995\)](#). Next, it holds

$$(C.10) \quad b^{-2}(\Psi^* - \Psi) = \frac{b^{-2}}{\widehat{p}_1}((\theta_1^* - \theta_1) - (\theta_2^* - \theta_2)) + b^{-2}(\theta_1 - \theta_2)\left(\frac{1}{\widehat{p}_1} - \frac{1}{p_1}\right).$$

The second term on the right-hand side of (C.10) is equal to  $o_p(1)$  when  $b$  is of order  $O(n^{-1/(q+5)})$  or  $o(n^{-1/(q+5)})$ . Proposition A.2 ii) follows with Theorem 1 b) in [Nielsen and Linton \(1995\)](#). Finally, proposition A.2 iii) follows directly from the adapted proof of Theorem 1 c) [Nielsen and Linton \(1995\)](#) and the continuous mapping theorem.  $\square$

### C.4. Proofs of propositions in section 5.

*Proof of proposition 5.1.* First, ignoring  $X$ , note that due to proposition 3.1, the equality (3.2) implies the equality

$$(C.11) \quad P(T(t) > s) = P(T(t') > s).$$

With assumptions A3 and A4, (C.11) is equivalent to

$$P(T > s | Z = t) = P(T > s | Z = t').$$

Then, under condition (4.3), the estimators of each of the two probabilities is normally distributed with a variance  $\sigma^2(t)$ .  $\square$

#### REFERENCES

- Abbring, J. H. and van den Berg, G. J. (2003). The non-parametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517.
- Abbring, J. H. and van den Berg, G. J. (2005). Social experiments and instrumental variables with duration outcomes. Tinbergen Institute Discussion Paper 05 - 047/3, Tinbergen Institute, The Netherlands.
- Andersen, P., Borgan, Ø., Gill, R., and Keiding, N. (1997). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer New York.
- Biewen, M., Fitzenberger, B., Osikominu, A., and Paul, M. (2014). The effectiveness of public-sponsored training revisited: the importance of data and methodological choices. *Journal of Labor Economics*, 32(4):837–897.
- Bijwaard, G. (2008). Instrumental variable estimation for duration data. Tinbergen Institute Discussion Paper 08-032/4, Tinbergen Institute, The Netherlands.
- Bijwaard, G. and Ridder, G. (2005). Correcting for selective compliance in a re-employment bonus experiment. *Journal of Econometrics*, 125:77–111.
- Blasco, S. (2009). Do people forgo extra money to avoid job search assistance? Discussion paper, CREST, Paris, France.
- Bloom, H. S. (1984). Estimating the effect of job-training programs, using longitudinal data: Ashenfelter's findings reconsidered. *The journal of human resources*, 19:544–556.
- Blundell, R., Dias, M. C., Meghir, C., and Reenen, J. (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, 2(4):569–606.

- Carney, T. and Ramia, G. (2011). Welfare support and sanctions for noncompliance in a recessionary world labour market: Post-Neoliberalism or not? Technical Report 11, Sydney Law School.
- Cattaneo, M. D., Jansson, M., and Ma, X. (2017). Simple local polynomial density estimators. Working paper, University of Michigan, Institute for the Study of Labor (IZA).
- Chesher, A. (2002). Semiparametric identification in duration models. CeMMAP working paper CWP20/02, Centre for Microdata Methods and Practice, London, UK.
- Crépon, B., Dejemeppe, M., and Gurgand, M. (2005). Counseling the unemployed: does it lower unemployment duration and recurrence? IZA Discussion paper 1796, Institute for the Study of Labor (IZA), Bonn, Germany.
- Crépon, B., Ferracci, M., and Fougère, D. (2012). Training the unemployed in France: how does it affect unemployment duration and recurrence? *Annales d'Economie et de Statistique*, (107-108):175–199.
- Crépon, B., Ferracci, M., Jolivet, G., and Van den Berg, G. J. (2009). Active labor market policy effects in a dynamic setting. *Journal of the European Economic Association*, 7(2-3):595–605.
- Crepon, B., Ferracci, M., Jolivet, G., and van den Berg, G. J. (2010). Analyzing the Anticipation of Treatments Using Data on Notification Dates. IZA Discussion Papers 5265, Institute for the Study of Labor (IZA).
- De Giorgi, G. (2005). The New Deal for Young People five years on. *Fiscal Studies*, 26(3):371–383.
- Debauche, E. and Jugnot, S. (2007). Les effets du projet d'action personnalisé sur les sorties des listes de l'ANPE. Working paper, DARES, Paris, France.
- Duflo, E., Glennerster, R., and Kremer, M. (2007). Chapter 61: Using randomization in development economics research: A toolkit. volume 4 of *Handbook of Development Economics*, pages 3895 – 3962. Elsevier.

- Eberwein, C., Ham, J. C., and LaLonde, R. J. (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(4):655–682.
- Field, E. (2007). Entitled to work: Urban property rights and labor supply in Peru. *The Quarterly Journal of Economics*, 122(4):1561–1602.
- Fougère, D., Kamionka, T., and Prieto, A. (2010). L'efficacité des mesures d'accompagnement sur le retour à l'emploi. *Revue Économique*, 61(3):599–612.
- Freyssinet, J. (2002). La réforme de l'indemnisation du chômage en France. IRES Document de Travail 02.01, IRES.
- Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (2007). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics*, 4:65–78.
- Hahn, J., Todd, P., and der Klaauw, W. V. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.
- Ham, J. C. and LaLonde, R. J. (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica*, 64(1):175–205.
- Hausman, J. A. and Woutersen, T. (2014). Estimating a semi-parametric duration model without specifying heterogeneity. *Journal of Econometrics*, 178:114 – 131. Annals Issue: Misspecification Test Methods in Econometrics.
- Hausman, J. A. and Woutersen, T. M. (2008). The proportional hazard model. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke.
- Heckman, J. J. (2008). Econometric causality. *International Statistical Review*, 76(1):1–27.
- Heckman, J. J., LaLonde, R. J., and Smith, J. A. (1999). The economics and econometrics of active labor market programs. In Ashenfelter, O. and Card, D., editors, *Handbook of*

- Labor Economics*, volume 3 of *Hanbook of Labor Economics*, chapter 31, pages 1865–2097. Elsevier.
- Heckman, J. J. and Navarro, S. (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, 136(2):341–396.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.
- Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, 64(4):555–574.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. John Wiley & Sons, New York.
- Karunamuni, R. J. and Alberts, T. (2005). A generalized reflection method of boundary correction in kernel density estimation. *Canadian Journal of Statistics*, 33(4):497–509.
- Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics*, 142(2):785–806.
- Lalive, R., van Ours, J. C., and Zweimller, J. (2005). The effect of benefit sanctions on the duration of unemployment. *Journal of the European Economic Association*, 3(6):1386–1417.
- Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47(4):939–956.
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press.
- Le Barbanchon, T. (2012). The effect of the potential duration of unemployment benefits on unemployment exists and match quality in France. Working paper, Crest.
- Meyer, B. D. (1996). What have we learned from the Illinois reemployment bonus experiment? *Journal of Labour Economics*, 14(1):26–51.
- Michael, L., Ruth, M., and Conny, W. (2011). Long-run effects of public sector sponsored training in west germany. *Journal of the European Economic Association*, 9(4):742–784.

- Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.
- Müller, H. G. and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 50(1):61–76.
- Nielsen, J. P. and Linton, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *The Annals of Statistics*, 23(5):1735–1748.
- Pearl, J. (2000). *Causality. Models, reasoning, and inference*. Cambridge University Press.
- Robins, J. M. and Tsiatis, A. A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics - Theory and Methods*, 20:2609–2631.
- Sianesi, B. (2004). An evaluation of the Swedish system of Active Labor Market Programs in the 1990s. *The Review of Economics and Statistics*, 86(1):133–155.
- Tchetgen, E. J., Walter, S., Martinussen, T., and Glymour, M. (2014). Instrumental variable estimation in a survival context. Biostatistics working paper series 179, Harvard.
- Todd, P. E. (2007). Chapter 60: Evaluating social programs with endogenous program placement and selection of the treated. volume 4 of *Handbook of Development Economics*, pages 3847 – 3894. Elsevier.
- Tsiatis, G. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci.* 72, 72(1):20–22.
- Van den Berg, G. J. (2001). Duration models: specification, identification, and multiple durations. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, chapter 55, pages 3381–3460. Elsevier.
- Van den Berg, G. J., Bozio, A., and Costa Dias, M. (2014). Policy discontinuity and duration outcomes. *Working Paper*.
- Vikström, J. (2014). IPW estimation and related estimators for evaluation of active labor market policies in a dynamic setting. Working Paper Series, Center for Labor Studies

2014:8, Uppsala University, Department of Economics.