UNIVERSITY OF MANNHEIM

NONPARAMETRIC INSTRUMENTAL VARIABLE METHODS FOR DYNAMIC TREATMENT EVALUATION

Gerard J. van den Berg   Petyo Bonev   Enno Mammen

NONPARAMETRIC INSTRUMENTAL VARIABLE METHODS FOR DYNAMIC TREATMENT EVALUATION

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

ABSTRACT. We develop a nonparametric instrumental variable approach for the estimation of average treatment effects on hazard rates and conditional survival probabilities, without model structure. We derive constructive identification proofs for average treatment effects under noncompliance and dynamic selection, exploiting instrumental variation taking place during ongoing spells. We derive asymptotic distributions of the corresponding estimators. This includes a detailed examination of noncompliance in a dynamic context. In an empirical application, we evaluate the French labor market policy reform PARE which abolished the dependence of unemployment insurance benefits on the elapsed unemployment duration and simultaneously introduced additional active labor market policy measures. The estimated effect of the reform on the survival function of the duration of unemployment duration is positive and significant. Neglecting selectivity leads to an underestimation of the effects in absolute terms.

1. Introduction

In the evaluation of treatment effects on duration outcomes, such as the effect of job search assistance on unemployment durations, it is often interesting to distinguish effect sizes by the elapsed duration of unemployment. 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 studies therefore tend to estimate effect sizes on hazard rates or on conditional survival probabilities at a range of elapsed durations.

However, the identification of such dynamic treatment effects is hampered by some hurdles even if the assignment is randomized. First, suppose the treatment is randomized at some elapsed duration t after inflow into some state of interest. In the presence of unobserved determinants of the outcome, their distributions among survivors at some later point in time will differ across different treatment arms; see Meyer (1996), Ham and LaLonde (1996), Eberwein et al. (1997) and Abbring and van den Berg (2005). A second hurdle is posed by the standard issue of noncompliance. If individuals can choose a treatment status different from the one that has been assigned to them, then estimation results will suffer from the standard selection bias. We refer to these two hurdles as dynamic and static endogeneity, respectively. A third hurdle is posed by the fact that duration variables are often subject to right-censoring. In this paper, we develop an instrumental variable (IV) approach for identification and estimation of dynamic treatment effects on the conditional survival function and the hazard of a duration variable. Our method solves the dynamic and static endogeneity problems and allows for right-censoring. We do not adopt parametric or semiparametric structures. We also 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. Our estimators are dynamic versions of the Wald estimator.

We focus on a setting in which a single comprehensive treatment is assigned at a specific calendar point in time to all individuals in some state of interest. A typical example is a labor market policy reform that changes the unemployment benefits system. Cohorts of individuals receive the treatment at the same point in calendar time but at different elapsed durations of their spells. The policy intervention can be regarded as exogenous, but due to dynamic selection the distribution of unobserved characteristics at the moment of treatment will differ across cohorts. Additionally, we allow for noncompliance in the sense that individuals may influence the extent to which they are exposed to the new policy regime. As an alternative example, we may replace the role of the labor market policy reform by a randomized field experiment.

Van den Berg et al. (2014) considered exogenous policy interventions and demonstrated that nonparametric causal inference of effects on hazard rates and conditional survival probabilities greatly benefits from the availability of data in which ongoing spells are interrupted by the intervention. In particular, such data allow for a comparison of subsamples of treated and not-yet treated that experienced the same dynamic selection pattern at durations before the elapsed duration at which the treated subsample was exposed to the

intervention. Our approach also exploits ongoing spells that are interrupted by an exogenous intervention. The major contribution of our paper is to allow for partial compliance. Thus, the treatment assignment is not mandatory, and only some of those assigned select into it. The problem of noncompliance has received much attention in the static evaluation literature in recent years (references are provided below). This contrasts to noncompliance in a dynamic nonparametric context. We achieve identification of treatment effects using the time to assigned treatment as an instrument for the actual treatment status. Notice that we effectively have a setting in which the instrumental variable and the treatment indicator are realized at the same elapsed duration. This serves to prevent that individuals respond to the instrumental variable before the treatment indicator is realized, in which case the dynamic selection pattern would differ between the subsamples of those who are assigned to the treatment and those who are not.

In the second part of the paper we evaluate the French 2001 labor market policy reform PARE which changed the dependence of unemployment benefits on the elapsed unemployment duration and simultaneously introduced additional active labor market policy 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. In this empirical analysis we apply the methods devised in the first part of the paper. This includes an extensive examination of the plausibility of the assumptions required for the use of the methods. We address the non-testable independent right-censoring assumption in a simulation study. This suggests that the estimation results are robust to violations of the assumption, primarily because violations that are likely to occur in the PARE setting have opposite directions and offset each other's impact on the estimates.

An additional contribution of our paper concerns the development of a theoretical framework to analyze the importance of endogeneity due to noncompliance in a dynamic setting. Specifically, we propose how to measure the extent of noncompliance and the bias that would be induced if its role is ignored. Understanding noncompliance is an important ingredient in the analysis of policy effectiveness and policy design. Pilot studies with noncompliance can be used to derive bounds for the effect of a comprehensive policy reform with perfect compliance. Our methods are based on a comparison of untreated noncompliers with a whole nontreated cohort at the same elapsed duration. In the empirical analysis, the results indicate that noncompliance is endogenous and that one major reason for noncompliance is the expectation of a quick exit. These findings are in line with Blasco (2009) who studied noncompliance in the PARE reform.

By dealing with both dynamic and static selection, our paper provides a link between the IV literature on treatment effects, the literature on dynamic treatment evaluation, and the regression discontinuity literature. The emphasis on noncompliance and IV estimation means that the link to the existing literature on IV in survival models and dynamic models is particularly strong. Much of the latter literature is surveyed in Abbring and van den Berg (2005). Eberwein et al. (1997) were the first to introduce IV in econometric survival analysis. They applied this to study the causal effect of training on unemployment

durations. See also Robins and Tsiatis (1991), Chesher (2002), Bijwaard and Ridder (2005), Heckman and Navarro (2007), Bijwaard (2008) and Tchetgen et al. (2014). Typically, these studies adopt a semiparametric or a parametric model structure.[1] Abbring and van den Berg (2005) develop a nonparametric IV estimator of the local average treatment effect on the survival function for the case that instrument and treatment indicator are realized at the inflow into the state of interest.

Another branch of literature that is relevant for our study comprises of existing empirical evaluations of the PARE reform. These impose semiparametric or parametric model structures and/or focus on other outcome measures than we do. They are discussed in section 3 below. The remainder of this paper is structured as follows. We present our IV approach in section 2. In section 3, we apply our IV method to the French labor market policy reform PARE. Section 4 concludes. All proofs are in the appendix.

2. Identification and estimation of dynamic treatment effects

2.1. **Notation and a framework for dynamic treatment evaluation.** Assume that all agents in some state of interest O are assigned to receive a treatment at a specific calendar point in time $r > 0$. We are interested in the causal effect of this treatment on the distribution of the duration of stay in O. We embed our analysis in a framework with dynamic potential outcomes. We assume that potential outcomes of the individual i depend on pretreatment characteristics X_i and V_i, of which the q-dimensional X_i is observed, $q \geq 1$, and the one-dimensional V_i not. Let the random variable Z_i denote the time from inflow to the assigned point in time of treatment and S_i the elapsed duration in O at which individual i actually receives the treatment. S_i is a choice variable whereas Z_i is exogenous. For each $X = x, V = v, Z = z, S = s$, denote with $T_i(s, z, x, v)$ the potential duration of stay in O of individual i if he or she had characteristics (x, v) and received (z, s) as values for (Z, S). We allow $T_i(s, z, x, v)$ to be a random variable. This assumption reflects some intrinsic uncertainty in the transition, not necessarily observed and/or controlled by the agent, see Lancaster (1990) for a discussion. Throughout the paper, we assume that Z is an exclusion restriction in the sense that $T_i(s, z, x, v) = T_i(s, x, v)$. For notational simplicity, we will suppress the dependence on X and V as well as the individual index i and write simply $T(s)$.

This setup corresponds to a labor market program implementation, in which a policy reform is administered at a fixed point in time. Our methods however, as shown in the discussion below, can be extended to a setup with ongoing programs, in which the treatment is assigned at random points in time to different individuals. In a labor market context, X might be education, gender, number of siblings, age and experience at inflow, whereas V might be the ability of an unemployed or his or her motivation. In a medical study, X might be some observed health marker, whereas V might be some genetic unobserved component. X and V obtain values in Ω_X and Ω_V.

---

[1]The use of dynamic discrete choice models such as the reduced-form model in Heckman and Navarro (2007) enables the evaluation of complex treatment effects as well as the distribution of counterfactuals. Identification allows for general time-varying unobservables but uses identification at infinity as well as some separability and random effects assumptions.

We enrich this dynamic framework by allowing the agents to opt out of the assigned treatment. We refer to this opting out as static selection. To fix ideas, for each $z \in \mathbb{R}_+$ and each $(x, v) \in \Omega_X \times \Omega_V$, let the random variable $S(z, x, v)$ denote the potential compliance status of an individual with observed and unobserved characteristics x and v, respectively, given that the treatment z is assigned to that individual. For notational simplicity, we write $S(z)$. $S(z)$ can be interpreted as the potential elapsed duration in O at which an agent would like to be treated, if he or she was assigned to be treated at elapsed duration z. To make the model tractable, an agent is only allowed to accept or reject an assigned treatment, and the treatment is only offered once (see assumption A1 in the following subsection, as well as the corresponding discussion). Thus, for each $z \in \mathbb{R}_+$, $S(z)$ may take only the values z (the case of compliance) and ∞ (the case of noncompliance).[2] Agents are allowed to have an arbitrary time structure of their compliance preferences. A cancer suffering patient might be reluctant to accept a new therapy at an early stage of the disease, but his or her preference might change at an advanced stage of the disease. Similarly, an unemployed person might refuse a training early in the unemployment spell and be willing to attend it later on. To account for the possibility of changing preferences, we refer to individuals who would be willing to receive a treatment at some elapsed duration z, given that they were asked to do so, as z-compliers. This notion generalizes the static compliance definition.

Allowing for static selection is common in the standard literature on (static) treatment evaluation, see Heckman and Vytlacil (2007). In a labor market program, unemployed individuals might decide not to accept an offer for a training or a counselling service. In a medical study, patients assigned to drop out from a therapy might be able to participate in a substitute program. Selection into or out of a certain treatment status creates a potential endogeneity problem, which has given rise to the development of the Local Average Treatment Effect (LATE) literature, see Imbens and Angrist (1994). Typically, the randomized treatment assignment is used as an instrument for the endogenous actual treatment status.[3]

Let T be the actual duration of the spell. T might be right censored by a random variable C. Define $\widetilde{T} := \min\{T, C\}$ and the censoring indicator $\delta := 1\{\widetilde{T} = T\}$. We observe $(\widetilde{T}, \delta)$ and not directly (T, C). We assume access to an i.i.d. sample

$$(\widetilde{T}_1, S_1, Z_1, X_1, \delta_1), \ldots, (\widetilde{T}_n, S_n, Z_n, X_n, \delta_n),$$

where S_i is missing if $S_i > \widetilde{T}_i$.

Remark

Unless explicitly otherwise stated, we will denote with t, s, z elapsed durations in O (and not calendar time). Thus, for example, 0 refers to the point in time of inflow of an agent into O. Furthermore, we do not need a binary process $D_i(t)$ that denotes the treatment

---

[2]Alternatively, we might restrict the maximal potential duration of the state of interest to be equal to some positive real number $\bar{S}$. In that case, noncompliers receive $S(z) = \bar{S}$. We do not differentiate between these two cases and write ∞.

[3]In line with the biometry literature, this instrument is also called Intention-to-Treat (ITT)

status of an agent i at time t. Before the calendar point in time r, nobody is treated. After r, all compliers are treated, that is, all individuals whose value of S is equal to the corresponding value of Z. Therefore, the treatment status can be deduced from S, Z and the calendar time.

Let $t \geq t''$. The treatment effect of interest is

$$(2.1) \qquad P(T(s) \in [t, t+a] \mid T(s) \geq t'', X, V) - P(T(s') \in [t, t+a] \mid T(s') \geq t'', X, V),$$

that is, the additive effect of replacing the treatment s' with the treatment s on the probability to exit the state of interest between t and $t + a$ conditionally on surviving up to t''. The case $s' = \infty$ induces a comparison between those treated at s and those never treated. Another special case is the limit case $a \to 0$, $t'' = t$. Denote with $\theta_{T(s)}(t \mid X, V)$ the hazard of $T(s)$ at t for an individual with characteristics X and V. Then the individual additive treatment on the hazard at t is defined as

$$(2.2) \qquad \theta_{T(s)}(t \mid X, V) - \theta_{T(s')}(t \mid X, V).$$

It reflects the additive change in the exit rate induced by a change of the treatment from s' to s. One appealing feature of additive treatment effects is their intuitive interpretation. To see this, write $P(T(s) \in [t, t+a] \mid T(s) \geq t'') = \mathbb{E}[1\{T(s) \in [t, t+a]\} \mid T(s) \geq t'', X, V]$. The indicator function is a Bernoulli random variable and its distribution is completely determined by its expectation.

One might be interested in identifying the (additive) effect on the unconditional survival function, that is, $t'' = 0$:

$$(2.3) \qquad P(T(s) \in [t, t+a)) - P(T(s') \in [t, t+a)),$$

However, this precludes dynamic selection; see Abbring and van den Berg (2005) for a discussion.[4] Often though it might be of interest to identify the effect of a treatment assigned at a later point in time only for those who actually would receive the treatment. In the labor market example, such a case would arise if a treatment is targeted at longterm unemployed individuals. In the medical example, due to its side effects, a therapy might be targeted only at patients who are at an advanced stage of a disease. For this reason, we consider the general case of conditioning on survival up to a point $t'' = t$ for $0 \leq t = s < s' \leq \infty$, that is

$$(2.4) \qquad P(T(t) \in [t, t+a] \mid T(t) \geq t, X, V) - P(T(s') \in [t, t+a] \mid T(s') \geq t, X, V).$$

Conditioning on survival has one further justification. Note that allowing for noncompliance requires the observability of the compliance status. In our framework, individuals who exit the state of interest prior to revealing compliance preferences have an unknown compliance status (that is, we do not know whether they are compliers).

We do not impose a parametric form on the distribution of $T(s)$ and we allow for separability and general dependence of observed and unobserved covariates X and V, respectively. The restriction $t = s$ is necessary to "unify" the dynamic selection between treated and untreated, as discussed in the next subsection. By redefining s to be the time

---

[4]Abbring and van den Berg (2005) consider a case with conditioning on a positive elapsed spell duration, $t'' > 0$, that is, conditioning on $T(s) > t''$, $t'' > 0$, and derive bounds for the effect.

to dropout of a treatment, we can analyze the effect of the length s of a treatment on the distribution of $T(s)$.

There are two limitations we have to consider. First, not specifying the dependence of the distributions of $T(s)$ and the unobservables V makes it impossible to identify the individual treatment effect (2.4). The price to pay for the functional form generality is that we have to average V out. Due to dynamic selection, the distribution of the unobservables might 1) be different in the subpopulation of survivors at some point in time $t > 0$ from the distribution in the whole population and 2) differ among different treatment arms. Therefore, it arises the question over which distribution of V to average. Van den Berg et al. (2014) suggest to condition on different subpopulations of survivors, such as treated survivors. A second limitation that arises in our context due to the possibility of static selection is that one can observe only the t-compliers with the treatment. This problem has been discussed in the literature on static treatment effects, see Imbens and Angrist (1994). Their solution is to consider only a treatment effect on the subpopulation of compliers. We adapt this restriction to our dynamic concept of compliance. We condition on $S(t) = t$. This restricts the analysis to the subpopulation of t-compliers, that is, to those individuals who would take the treatment at an elapsed duration of t if they were asked to do so. With these considerations, we define the Average Treatment Effect on the Treated Complying Survivors, shortly TE, as

$$(2.5) \qquad TE(t, t', a) := \mathbb{E}[P(T(t) \in [t, t+a] \mid T(t) \geq t, S(t) = t, X, V) -$$
$$P(T(t') \in [t, t+a] \mid T(t') \geq t, S(t) = t, X, V) \mid T(t) \geq t, S(t) = t, X].$$

The effects on the nontreated and on the whole population are defined analogously.[5] The positive constant a is chosen such that $a < t' - t$. This restriction insures a comparison of treated with nontreated individuals. Similarly, the treatment effect on the hazard (HTE) is defined as

$$(2.6) \qquad HTE(t, t') := \mathbb{E}[\theta_{T(t)}(t \mid S(t) = t, X, V) -$$
$$\theta_{T(t')}(t \mid S(t) = t, X, V) \mid T(t) \geq t, S(t) = t, X].$$

Remark

An alternative treatment effect that can be considered in this framework is a relative effect on the hazard rate at t, $\theta_{T(s)}(t \mid X, V)/\theta_{T(s')}(t \mid X, V)$. Abbring and van den Berg (2005) prove identification of this treatment effect under multiplicative unobserved heterogeneity, that is, under $\theta_{T(s)}(t \mid X, V) = \theta^*_{T(s)}(t \mid X).V$. We do not pursue this approach here.

Remark

Our model can be applied to an alternative setup, in which individual spells have the same starting point 0 in calendar time, but the agents receive the treatment at different points in time. Here, a cohort $\{Z = t\}$ consist of all individuals who are assigned to receive the treatment at calendar time t.

---

[5]In fact, they coincide under the assumptions introduced in the next subsection, see proposition 2.1.

2.2. **Identification of dynamic treatment effects.** In this section, we show that there exists a function that links the joint distribution of the observables with the treatment effect. As a result, the treatment effect is identified. We derive this function explicitly. Thus, our identification strategy is constructive in the sense that it provide a guidance for estimation. We adopt the following assumptions:

A1 **(Single treatment)** : for any t it holds either $S(t) = t$ or $S(t) = +\infty$.

A2 **(No anticipation)** : For each real $t' \geq t \geq 0$ and each X, V holds

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

where $\Theta_{T(s)}$ is the integrated hazard of $T(s)$.

A3 **(Randomization)** : For the instrument Z it holds

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

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

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

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

(1) Assumption A1 defines the possible types 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 $t' < \infty$. A1 is compatible with a setup where the treatment is administered at a single point in calendar time and agents have no access to an alternative treatment. This setup corresponds to a one-sided noncompliance in the static treatment evaluation literature. One-sided noncompliance precludes the existence of always-takers.[6] As a result, no monotonicity-type assumption (as the one invoked in Imbens and Angrist (1994)) is needed for identification. 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 ∞.

(2) Assumption A2 states basically that future treatments are not allowed to influence the past. The assumption implies that the individual probability of a survival up to t is the same for any two future treatments $t', t'', t \leq t', t''$. In a model with forward looking agents, A2 requires that agents either have no knowledge on the point in time of treatment (i.e. they do not anticipate it) or that they do not act upon that knowledge. Technically, jointly with assumption A3, the "no anticipation" assumption is used to ensure equal pretreatment patterns of dynamic selection in the different treatment arms, see proposition (2.1) below as well as the discussion in the paragraphs right before and after proposition (2.1). In the context of active labor market policies, the "no anticipation" assumption is plausible in numerous settings (see Abbring and van den Berg (2003) for a discussion). Often the start of a training program and the assignment to treatment are dictated by budget and other administrative reasons and appear to the unemployed as random. Those assigned to the treatment might be chosen at random from all eligible unemployed.

---

[6]See Imbens and Angrist (1994) for the definitions of always-takers and never-takers.

Moreover, the assignment may occur without a preliminary notice so that the timing is unexpected to the unemployed. This is almost by definition true for punitive treatments such as sanctions. Notice also that the exact content and point in time of implementation of a policy reform are often a subject to persistent debates. The resulting uncertainty might deter agents from building an anticipation about start and content of the reform. Most of the empirical evaluation literature on active labor market policies tacitly assumes absence of anticipatory effects. In our empirical application, we argue in detail that this extends to the case of the French policy reform PARE. Note that conditioning on $S(t), S(t')$ ensures that we adopt the "no anticipation" assumption for all relevant subpopulations, in particular for the subpopulation of compliers, $\{S(t) = t\}$, and for the subpopulation of noncompliers, $\{S(t) = \infty\}$.

(3) Assumption A3 is a randomization assumption. A3 i) implies that once we condition on observables and unobservables, there is no selection into the different treatment assignments. Taken together, i) and ii) imply the conditional independence assumption

(2.7) $$Z \perp\!\!\!\perp \{T(s), S(t)\} \mid X.$$

In the empirical analysis, A3 requires a stable (macro-) economic environment in the period of consideration. Economic structural brakes and mass layoffs might cause a violation of A3. A version of the implication (2.7) is testable, see the empirical investigation below for a discussion.

(4) The consistency assumption implies that a potential outcome corresponding to a given treatment is observed if the treatment is actually assigned. Another way to write it is $T = T(S), S = S(Z)$. A4 provides the link between potential outcomes and observations and is necessary for identification.

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 imply either that S and Z are discrete or that at least they have a positive probability mass on t and t'.[7] 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 counselling 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.

Suppose for the moment that T is observable (the case with right censoring is considered at the end of this subsection). As a motivation for our identification strategy, consider first

---

[7]If $a = b = 0$, then we define the expression a/b to be equal to 0.

the following naive candidates for a treatment effect:

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

and

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

for $t' > t$. Writing (2.8) in the form

$$\mathbb{E}[P(T \in [t, t+a) \mid T \geq t, X, S = t, Z = t, V) \mid T \geq t, X, S = t, Z = t] -$$
$$\mathbb{E}[P(T \in [t, t+a) \mid T \geq t, X, S = \infty, Z = t, V) \mid T \geq t, X, S = \infty, Z = t]$$

makes it clear that it compares averages over two different subpopulations of the same cohort: the t-compliers and the t-noncompliers. These two subpopulations might have different distributions of the unobserved heterogeneity V because the treatment status S is a choice variable. As a consequence, it would hold that $V \not\perp S \mid T \geq t, X, Z = t$. We will refer to this consequence as static endogeneity or static selection. The (potential) endogeneity arises immediately with the decision to accept or refuse the treatment. As a result, (2.8) would capture not only the treatment effect but also the bias from the static selection. We use the naive treatment effect (2.8) to analyze the nature of endogeneity. We compare it to our IV estimator to construct a test for exogeneity, see section 2.5 for details. The difference between (2.8) and the IV estimator is informative about the selection process. A better understanding of the selection might be used to impose more structure on the model. In our empirical application, an estimator of (2.8) is shown to underestimate the positive treatment effect. Hence, the control group must contain many quick exits, which sheds light on the reasons for the non-take up of the reform.

The naive treatment effect (2.9) compares the average outcome of the t-compliers from the younger cohort $\{Z = t\}$ with the average outcome of the t'-compliers of the older cohort $\{Z = t'\}$. Due to dynamic selection, this comparison amounts to averaging over two potentially different distributions of V. (2.9) can be used to shed light on the nature of this dynamic selection process.

Both examples demonstrate the importance and difficulty of the choice of a treatment and a control groups in a setting with static and dynamic selection. We propose a strategy that can deal with both types of selection. The intuition for this strategy is as follows. An appealing choice for a treatment group is the set of compliers from the cohort $\{Z = t\}$: consistency links observed outcomes of the treated compliers with the potential outcomes. Suppose for the moment that we observe the potential compliance status at any point in time. Then, one possible control group for the treated t-compliers from cohort $\{Z = t\}$ would be the not yet treated group of t-compliers from the older cohort who survive at least t time units. The intuition behind this choice is the following. If the unobserved heterogeneity V has the same distribution in the two cohorts at the point in time of inflow, 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 pretreatment elapsed duration $t > 0$. The equality of the distributions of V at $t = 0$ is ensured by the randomization assumptions A3 i) and ii). The dynamics is controlled by the "no anticipation" assumption A2. This idea is first

developed in Van den Berg et al. (2014) for the case of perfect compliance. It amounts to a direct comparison of the average outcomes of two cohorts. In a first step, we generalize the result of Van den Berg et al. (2014) to a setting with endogenous compliance.

Proposition 2.1. *Let F be a cdf. Under assumptions A1 to A4, it holds for all $\infty \geq t' \geq t \geq 0$*

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

Proposition 2.1 states that the unobservables have the same dynamics for two potential treatments on the set of t-compliers. It also links the distribution of V given a potential treatment to the distribution of V in the subpopulation of observed t-compliers, $\{S = t, Z = t\}$. There are two immediate consequences of proposition 2.1. First, the treatment effects on the treated survivors, on the nontreated survivors and on all survivors, respectively, coincide. Second, the following result holds:

Corollary 2.1. *Let $a \leq t' - t$. Under Assumptions A1-A4, it holds for all $\infty \geq t' \geq t \geq 0$*

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

The following proposition provides the key identification result:

Proposition 2.2. *Let $a \leq t' - t$. Under Assumptions A1-A4, $TE(t, t', a)$ is nonparametrically identified for all $\infty \geq t' \geq t \geq 0$ and it holds*

(2.10)

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

The intuition for identification is the following. Corollary 2.1 provides a direct hint on how to choose the treatment group. t-compliers from the cohort $\{Z = t\}$ reveal their preferences at the point in time of treatment. We can therefore link potential and observed outcomes using A4, proposition 2.1 and corollary 2.1. The main obstacle for constructing a control group is that we do not observe the compliance status of individuals in the older cohort $\{Z = t'\}$ at elapsed duration t. Agents reveal their preferences at the time of treatment. In line with the argumentation above, due to dynamic selection, the sub-population of t'-compliers differs from the subpopulation of t-compliers in terms of the distribution of V. The key to identification is the observation, that the potential outcome corresponding to a certain treatment is the sum of potential outcomes of compliers and noncompliers, weighted by their proportions. Written in a simplified notation, we have

(2.11) $F_0 = F_{C,0}P_C + F_{N,0}P_N,$

where the zero indicates the no-treatment case[8], and, with a temporary abuse of notation, C and N denote compliers and noncompliers, respectively. In order to link $F_{C,0} = (F_0 - F_{N,0}P_N)/P_C$ to observables, it is sufficient to express F_0, P_C, P_N and $F_{N,0}$ in terms of observables. Due to assumptions A1-A4 (in particular to no anticipation), the average

---

[8]The correct expression should be "not yet treated-case". Under the "no anticipation" assumption, however, this distinction does not matter in the interval $[t, t + a)$.

outcome $F_{N,0}$ of the noncompliers of the older cohort $\{Z = t'\}$ is equal to the average outcome of the noncompliers from the treatment group $\{Z = t\}$, see Lemma A.1, part 2. Both subgroups do not anticipate and do not receive the treatment. The average outcome of the noncompliers from $\{Z = t\}$ is identified, see Lemma A.1, part 1. Note that an important implication of randomisation, no anticipation and consistency is that the actual assignment of the treatment does not change the behaviour of noncompliers. The proportions P_C and P_C are identified in an analogous way, see lemma A.2. Finally, F_0 can be linked directly to the outcomes of cohort $\{Z = t'\}$ and is also identified.

Expression (2.10) has an intuitive interpretation. It adjusts the difference between the average observed outcomes in the two cohorts by the probability to be a complier. The adjustment takes account of the fact, that any difference between the two cohorts can be caused only by the compliers. Our result is in the spirit of the static one-sided noncompliance result of Bloom (1984). This resemblance seems natural in a setting where agents are allowed to refuse an assigned treatment but are not able to select into an alternative treatment arm (i.e. choose a different point in time of treatment).

Unlike in the static treatment evaluation models, randomization alone is not enough to ensure identification. An experiment might be randomized at $t = 0$ but due to dynamic selection endogeneity arises over time. The "no anticipation" assumption precludes this possibility.

Remark

A special case of proposition 2.2 is the limit case $a \to 0$. We devote a separate section on its identification and estimation because of the importance and specifics of hazards.

Remark

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 $a \to 0$ simply for $t', t'' > t$). To see this, note that under A2, it holds $P(T(t') \in [t, t+a] \mid T(t') \geq t, X) = P(T(t'') \in [t, t+a] \mid T(t'') \geq t, X)$ for all $t', t'' \geq t+a$. On the other hand, the treatment effect TE (HTE) is identified only for t' that fulfils $t' \geq t + a$ (or $t', t'' > t$). As a consequence, it follows that the treatment effects do 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 $HTE(t)$.

Thus far, we have assumed we can observe the whole length of spells in the state of interest, T. A typical feature of duration data is that observations might be censored. In this paper, we consider right censoring.[9] In labor market studies, right censoring typically arises when at the end of the study the individuals are still unemployed, so the unemployment spell has an unknown length. The unemployed might also simply stop attending the training and drop out of the study (sample attrition). In addition, the job search might be interrupted by a transition out of the labor force due to maternity, sickness, military service or other reasons. In biomedical studies, and particularly in clinical trials, spells are right-censored when patients die from another cause (competing

---

[9]Extensions to left or interval censoring are straightforward.

risks) or withdraw from treatment. We introduce formally right censoring in the following way: let C be a real nonnegative random variable. We observe $(\widetilde{T}, \delta)$ and not directly (T, C). It is not possible to recover nonparametrically the joint distribution of T and C from the distribution of $(\widetilde{T}, \delta)$ without additional assumptions. The reason for this impossibility is a nonidentification result that goes back to Cox (1962) and Tsiatis (1975), namely that to each pair of latent variables (T_d, C_d) there exists an independent pair of variables (T_i, C_i) that is observationally equivalent to (T_d, C_d). To achieve identification, we adopt the following additional standard assumption:

A5) (Random censoring)

$$C \perp\!\!\!\perp (T, S) \mid X, Z.$$

Assumption A5 is nontestable due to the nonidentification result of Tsiatis (1975). In the context of our empirical application, we show with a Monte Carlo simulation that plausible violations of A5 offset each other. Thus, our estimation results are likely to be robust against violations of A5. With A5, we can prove the following proposition:

Proposition 2.3. *Under assumptions **A1 - A5** TE(t,a) is identified.*

The proof of proposition 2.3 is straightforward. The probabilities $P(T \in [t, t+a) \mid T \geq t, X, Z = j)$ for $j \in \{t, t'\}$ can be written as differences of survival functions and be estimated consistently with a Kaplan-Meier estimator, see section 2.3 for details. Note also that for the cohort $\{Z = t\}$, S is observed whenever $\widetilde{T} \geq t$,[10] so that due to A5

$$(2.12) \qquad P(S = t \mid T \geq t, X, Z = t) = P(S = t \mid T \geq t, C \geq t, X, Z = t) = P(S = t \mid \widetilde{T} \geq t, X, Z = t).$$

The last probability in (2.12) contains only observables and can be consistently estimated from the data.

2.3. **IV estimation of dynamic treatment effects.** To ease notation, probability and survival functions concerning the cohorts $\{Z = t\}$ and $\{Z = t'\}$ are denoted with an index 1 and 2, respectively. For example, we write $P_1(T \in [t, t+a) \mid T \geq t)$ instead of $P(T \in [t, t+a) \mid T \geq t, Z = t)$. Furthermore, we ignore the dependence on observed covariates X.[11] Assumptions A2 and A3 are adapted accordingly. The generalization to the case with covariates is straightforward. Denote with $\bar{F}_1$ and $\bar{F}_2$ the survival functions of T in the two cohorts, $\bar{F}_i(t) := P_i(T > t)$. A starting point for our estimation procedure is the equality

$$(2.13) \qquad TE(t,a) = \frac{1}{P_1(S = t \mid T \geq t)} \left(\frac{\bar{F}_2(t+a)}{\bar{F}_2(t)} - \frac{\bar{F}_1(t+a)}{\bar{F}_1(t)} \right),$$

which holds under assumptions A1-A4. It follows from the result in proposition 2.2 together with $P_i(T \in [t, t+a) \mid T \geq t) = 1 - \bar{F}_i(t+a)/\bar{F}_i(t)$. T might be censored so that we only observe $(\widetilde{T}, \delta)$. $\bar{F}_i(t)$ can be consistently estimated with the Kaplan-Meier estimator. Under

---

[10]This follows from assumptions A1 and A4.

[11]There are four cases of interest: estimation of TE with/without covariates and estimation of HTE with/without covariates. We consider two complementary cases, namely TE without covariates and HTE with covariates. The case TE with covariates follows in a straightforward way when one uses the estimator of Gonzalez-Manteiga and Cadarso-Suarez (2007) instead of the unconditional Kaplan-Meier estimator.

the independent censoring assumption A5 and additional mild regularity conditions, it holds

$$(2.14) \qquad \widehat{\bar{F}}_i(t) = \bar{F}_i(t) + o_p(1) \qquad \text{and}$$

$$(2.15) \qquad \sqrt{n}\left(\widehat{\bar{F}}_i(t) - \bar{F}_i(t)\right) \overset{d}{\to} N(0, \sigma_i^2(t)) \qquad \text{as} \quad n \to \infty,$$

where $\sigma_i^2(t)$ is the asymptotic variance of the Kaplan-Meier estimator, $t \in [0, \infty)$, see e.g. page 18 ff. Kalbfleisch and Prentice (2002).[12] The additional regularity conditions 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. All results hold for continuous as well as discrete time.

Next, under the independent right-censoring assumption, it holds

$$(2.16) \qquad P_1(S = t \mid T \geq t) = P(S = t \mid T \geq t, Z = t, C \geq t) = P_1(S = t \mid \widetilde{T} \geq t) =: p > 0.$$

p contains only observables and is nonparametrically identified. Let $\widehat{p} := \widehat{P}_1(S = t \mid \widetilde{T} \geq t)$ be a consistent nonparametric estimator of p. We define the IV-estimator $\widehat{TE}(t, a)$ of $TE(t, a)$ as

$$(2.17) \qquad \widehat{TE}(t, a) = \frac{1}{\widehat{p}}\left(\frac{\widehat{\bar{F}_2}(t + a)}{\widehat{\bar{F}_2}(t)} - \frac{\widehat{\bar{F}_1}(t + a)}{\widehat{\bar{F}_1}(t)}\right).$$

Its asymptotic properties are similar to those derived in Abbring and van den Berg (2005) for the case where both the instrument and the treatment status are realized upon inflow into the state of interest (i.e., $t = 0$). Here we allow for conditioning on survival up to t and we only consider one-sided noncompliance. The following proposition states the consistency of (2.17).

Proposition 2.4. *Suppose* (2.14) *holds. Then, under assumptions A1-A5, it holds*

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

for each admissible pair (t, a).

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 $\widehat{p}$.

Consider the Null hypothesis

$$(2.18) \qquad H_0: \qquad \text{(Ineffective treatment)} \qquad \frac{\bar{F}_2(t + a)}{\bar{F}_2(t)} - \frac{\bar{F}_1(t + a)}{\bar{F}_1(t)} = 0.$$

Under (2.18), it holds

$$\sqrt{n}\widehat{TE}(t, a) = \frac{\sqrt{n}}{\widehat{p}}\left(\frac{\widehat{\bar{F}_2}(t + a)}{\widehat{\bar{F}_2}(t)} - \frac{\widehat{\bar{F}_1}(t + a)}{\widehat{\bar{F}_1}(t)}\right) =$$

$$= \frac{\sqrt{n}}{\widehat{p}}\left(\frac{\widehat{\bar{F}_2}(t + a)}{\widehat{\bar{F}_2}(t)} - \frac{\bar{F}_2(t + a)}{\bar{F}_2(t)}\right) - \frac{\sqrt{n}}{\widehat{p}}\left(\frac{\widehat{\bar{F}_1}(t + a)}{\widehat{\bar{F}_1}(t)} - \frac{\bar{F}_1(t + a)}{\bar{F}_1(t)}\right)$$

---

[12]Equation (2.14) follows from equation (2.15).

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

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

and therefore

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

The last term converges to zero in probability.

With (2.15), the terms $\quad \frac{\sqrt{n}}{\bar{F}_i(t)}(\widehat{\bar{F}_i}(t+a) - \bar{F}_i(t+a)) \quad$ and $\quad \frac{\bar{F}_i(t+a)\sqrt{n}}{\bar{F}_i^2(t)}(\widehat{\bar{F}_i}(t) - \bar{F}_i(t))$

are asymptotically normally distributed with mean 0 and variances

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

With the independence of the random variables D_1 and D_2, where $D_i = \frac{\widehat{\bar{F}_i}(t+a)}{\widehat{\bar{F}_i}(t)}$, $i = 1, 2$, we can now state the following proposition.

Proposition 2.5. *Let assumptions A1-A5 and condition* (2.15) *hold. Then, under the null* (2.18), *it holds*

$$(2.19) \quad \sqrt{n}\widehat{TE}(t, a) \xrightarrow{d} N\left(0, \frac{1}{p^2}\sum_{i=1}^{2}\left(\frac{1}{\bar{F}_i^2(t)}\sigma_i(t+a) + \frac{\bar{F}_i^2(t+a)}{\bar{F}_i^4(t)}\sigma_i(t) + \frac{\bar{F}_i(t+a)}{\bar{F}_i^3(t)}\sigma_i(t, t+a)\right)\right),$$

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

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 (2.19) that the precision of the estimator is inversely related to p. The bigger the compliance probability p, i.e. the stronger the instrument Z for the endogenous S, the smaller the variance of the IV-estimator. This intuitive result is in line with the standard static IV literature. (2.17) can be interpreted as a dynamic version of the Wald estimator. A generalization to the case of covariates can be achieved by replacing the unconditional Kaplan-Meier estimator with the conditional estimator of Gonzalez-Manteiga and Cadarso-Suarez (2007), following the same steps as here.

2.4. **Identification and estimation of additive treatment effects on the hazard.** In this subsection, we state conditions under which the treatment effect on the hazard, (2.6), is identified and 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.

2.4.1. *Identification.* Write $W = (X, V)$ and let Ω_W be the set of possible values for W. Further, write $\Psi(t \mid X) := HTE(t, X)$ (we stress explicitly the dependence on X) and define $\theta(t \mid X) := \lim_{dt \to 0} P(T \in [t, t + dt \mid T \geq t, X))/dt$ (all expressions are assumed to exist). The rest of the notation is the same as in the last sections. Again we assume access to an i.i.d. sample

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

Our first result is the following

Proposition 2.6. *Let the measurable function* $g : \mathbb{R}^+ \times \Omega_W \to \mathbb{R}^+$ *fulfill* $\mathbb{E}[g(t, W)] < \infty$ *and* $\mid \theta(t \mid W = w) \mid \leq g(t, w)$ *for each* $(t, w) \in \mathbb{R}^+ \times \Omega_W$. *Then, under assumptions A1-A5,* $\Psi(t \mid X)$ *is identified and it holds*

$$(2.20) \qquad \Psi(t \mid X) := \frac{\theta(t \mid X, Z = t) - \theta(t \mid X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)}.$$

Under the Lebesque dominated convergence theorem,

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

and the proof follows directly from proposition 2.2. Thus, as expected, the *HTE* is revealed to be the limit case of the general treatment effect *TE*, $HTE = \lim_{dt \to 0} TE/dt$. In the case of a full compliance, that is $P(S = t \mid T \leq t, X, Z = t) = 1$, *HTE* reduces to $\theta(t \mid X, Z = t) - \theta(t \mid X, Z = t')$ which is the result of Van den Berg et al. (2014).

2.4.2. *Estimation.* Henceforth, we denote with $\theta_1(t \mid X)$ the hazard $\theta(t \mid X, Z = t)$ of the younger cohort, $\{Z = t\}$, and with $\theta_2(t \mid X)$ the hazard $\theta(t \mid X, Z = t')$ 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 $\Psi(t \mid X)$, only the observations $\widetilde{T}$ that are bigger than or equal to t are informative about $\theta_1(t \mid X)$.[13] This leads to estimating a hazard at the left boundary of the interval $[t, \bar{T})$ where $\bar{T}$ is some maximum duration, possibly ∞. 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 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

---

[13]This does not apply to $\theta_2(t \mid X)$.

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 $\Psi(t \mid x)$ at an interior point x of Ω_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 $K(x) = \Pi_{i=1}^{q} k(x(i))$, where $x = (x(1), \ldots, x(q))$. Next, let k_+ denote the asymmetric kernel function

$$k_+ : [0, 1] \times [-1, 1] \to \mathbb{R}$$
$$(h, y) \to \frac{12}{(1+h)^4}(y+1)[y(1-2h) + (3h^2 - 2h + 1)/2].$$

This is a boundary kernel function as defined in Müller and Wang (1994). The support of $k_+(h, .)$ is $[-1, h]$. In analogy to the symmetric kernel k, we define the second moments of $k_+(0, .)$ 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 $i = 1, \ldots, n$ the observed failure process of the ith individual at time t, $N_i(t) := 1\{\widetilde{T}_i \le t, T_i \le C_i\}$ and the individual process at risk, $Y_i(t) := 1\{\widetilde{T}_i \ge t\}$. 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, $X_{1,i}$ denotes an observation of X that comes from the cohort $\{Z = t\}$. Then our estimator $\widehat{\Psi}(t \mid x)$ of $\Psi(t \mid x)$ is defined as

$$(2.21) \qquad \widehat{\Psi}(t \mid x) := \frac{1}{\hat{p}_1(t \mid x)} \left(\frac{\sum_{i=1}^{n} K(\frac{x-X_{1,i}}{b}) \int k_+(t, \frac{t-s}{b})dN_{1,i}(s)}{\sum_{i=1}^{n} K(\frac{x-X_{1,i}}{b}) \int k_+(t, \frac{t-s}{b})Y_{1,i}(s)ds} \right.$$
$$\left. - \frac{\sum_{i=1}^{n} K(\frac{x-X_{2,i}}{b}) \int k_+(t, \frac{t-s}{b})dN_{2,i}(s)}{\sum_{i=1}^{n} K(\frac{x-X_{2,i}}{b}) \int k_+(t, \frac{t-s}{b})Y_{2,i}(s)ds} \right),$$

where $\hat{p}_1(t \mid x)$ is a nonparametric estimator for $p_1(t \mid x) := P(S = t \mid T \ge t, X = x, Z = t)$. We assume that $\hat{p}_1(t \mid x)$ is consistent. In addition, for proposition 2.7 ii) we assume that $b^{-2}(\hat{p}_1(t \mid x) - p_1(t \mid x)) = o_p(1)$, which can be assured by assuming that $p_1(t \mid x)$ is sufficiently smooth in x. The term

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

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

$$(2.22) \qquad \theta_j^*(t \mid x) := \frac{\sum_{i=1}^n K(\frac{x-X_{j,i}}{b}) \int k_+(t, \frac{t-s}{b}) \theta_j(s \mid X_{j,i}) Y_{j,i}(s) ds}{\sum_{i=1}^n K(\frac{x-X_{j,i}}{b}) \int k_+(t, \frac{t-s}{b}) Y_{j,i}(s) ds} \qquad j = 1, 2$$

and

$$(2.23) \qquad \Psi^*(t \mid x) = \frac{1}{\hat{p}_1(t \mid x)} (\theta_1^*(t \mid x) - \theta_2^*(t \mid x)).$$

We need the following assumptions.

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

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

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

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

Proposition 2.7. *Define*

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

Under assumptions H1-H3, the following results hold:

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

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

$$b^{-2}(\Psi^*(0 \mid x_0) - \Psi(0 \mid x_0)) \xrightarrow{p} \sum_{i=1}^2 \frac{(-1)^{i+1} k_1^+}{f_i(x_0) u_i(0) p_1(0 \mid x_0)}$$

$$\left[\frac{\partial \theta_i(0 \mid x_0)}{\partial t} \frac{\partial (f_i(x_0) u_i(0))}{\partial t} + \frac{1}{2} \frac{\partial^2 \theta_i(0 \mid x_0)}{\partial t^2} f_i(x_0) u_i(0) + \right.$$

$$\left. \sum_{j=1}^q \left(\frac{\partial \theta_i(0 \mid x_0)}{\partial x(j)} \frac{\partial (f_i(x_0) u_i(0))}{\partial x(j)} + \frac{1}{2} \frac{\partial^2 \theta_i(0 \mid x_0)}{\partial x(j)^2} f_i(x_0) u_i(0) \right) \right]$$

     iii) *Finally, it also holds*

$$\hat{\sigma}_\Psi^2 := \frac{nb^{q+1}}{\hat{p}_1(0 \mid x_0)^2} \sum_{j=1}^2 \frac{\sum_{i=1}^n K^2(\frac{x_0-X_{j,i}}{b}) \int k_+^2(\frac{-s}{b}) dN_{j,i}(s)}{(\sum_{i=1}^n K(\frac{x_0-X_{j,i}}{b}) \int k_+(\frac{-s}{b}) Y_{j,i}(s) ds)^2} \xrightarrow{p}$$

$$\sigma_\Psi^2$$

     Result i) gives the asymptotic distribution of the estimator, ii) characterizes the bias and iii) provides the standard errors for confidence bounds around Ψ^*. If the bandwidth is chosen to be of $o(n^{-1/(q+5)})$, then the asymptotic bias is negligible and proposition 2.7 can be used to construct confidence bands for Ψ.

2.5. **Framework for the analysis of endogeneity.** Understanding the nature of selection is important for setting up and evaluating a policy reform. Often a comprehensive policy reform is preceded by a small scale pilot study that allows for noncompliance. Understanding the non-take up of the pilot study might help better design the reform and derive bounds for its effect under perfect compliance. Better understanding of the endogeneity reasons can be used to model explicitly the selection process in more complex (e.g. general equilibrium) models. We develop a framework for answering the following two questions:

i) Is there endogenous selection caused by the decision of the agents to accept or refuse the treatment?

ii) If yes, in which direction would be the bias caused by the endogenous selection?

Answering the first question requires a specification of the possible channels of (static) endogeneity. In our framework, there are two potential endogeneity channels. First, unobserved characteristics of the agents determine both potential outcomes and the potential compliance decision. Second, the potential outcome itself (that is, after "controlling" for observed and unobserved individual characteristics) might influence the potential compliance status. The first channel amounts to a violation of

$$(2.24) \qquad S(t) \perp\!\!\!\perp \{T(s)\} \mid X$$

and the second of

$$(2.25) \qquad S(t) \perp\!\!\!\perp \{T(s)\} \mid X, V$$

We preclude the possibility of a violation of (2.25): we assume that the only way the potential outcome might influence the decision $S(t)$ is that the agent might have a knowledge of $T(s)$ and use it in the decision process. This individual knowledge of the potential outcome (or its distribution) is unobserved by the econometrician. It is therefore included in V.[14] With these considerations, we define the following null hypothesis:

$$(2.26) \qquad H_0: \quad S(t) \perp\!\!\!\perp \{T(s)\} \mid X$$

For $B := [t, t+a)$, (2.26) implies the following relation:

$$(2.27) \quad \widetilde{H}_0: P(T(\infty) \in B \mid T(\infty) \geq t, X, S(t) = t) - P(T(\infty) \in B \mid T(\infty) \geq t, X, S(t) = \infty) = 0.$$

Using A1-A4 and following the steps in proof of proposition 2.2, we obtain the equivalent to (2.27) relation

$$(2.28) \qquad \widetilde{H}_0: \quad P(T \in B \mid T \geq t, X, Z = t') - P(T \in B \mid T \geq t, X, S = \infty, Z = t) = 0.$$

Intuitively, if there is no selection, then the average observed outcomes of nontreated compliers and noncompliers should be the same. As a result, the average observed outcome of the whole cohort $\{Z = t'\}$ under no treatment (the left-hand side of (2.28)) should be equal to the average observed outcome of the noncompliers from the cohort $\{Z = t\}$ (the right-hand side of (2.28)). Equation (2.28) contains only observables. Deriving the distribution of the test statistics for (2.28) follows precisely the same steps as for the

---

[14]We have to assume that the agent does not learn about the potential outcomes over time. With a time-varying V, we would lose identification.

null hypothesis (2.18). We omit it here. A simplified testing procedure would induce a comparison of survival functions. The corresponding null hypothesis is

$$(2.29) \qquad \widetilde{H}_0: \qquad P(T \geq t \mid X, Z = t') - P(T \geq t \mid X, S = \infty, Z = t) = 0.$$

A test statistics is constructed by replacing the theoretical probabilities with their Kaplan-Meier estimators.

To answer question ii), we can compare the (theoretical proper) treatment effect (2.18) to the naive treatment effect (2.8). Written in the simplified notation of section 2.2, this is a comparison of a) $F_{C,1} - F_{C,0}$ and b) $F_{C,1} - F_{N,0}$. This ad hoc approach can be justified with (2.27). Recall that $F_{C,0} = (F_0 - F_{N,0}P_N)/P_C$. Subtracting b) from a), we obtain

$$(F_{C,1} - (F_0 - F_{N,0}P_N)/P_C) - (F_{C,1} - F_{N,0}) =$$
$$(F_{N,0}P_C + F_{0,N}P_N - F_0)/P_C = (F_{N,0} - F_0)/P_C.$$

The nominator $(F_{N,0} - F_0)$ of the last expression is precisely the left-hand side of (2.27).

3. Empirical Application: the French PARE labor market reform in 2001

3.1. **The reform.** We combine the IV method we developed in section 2 with a unique empirical strategy to analyze the effect of a reform in the French unemployment insurance system on the duration of unemployment. The new system, called Plan d'Aide au Retour á l'Emploi (PARE hereafter), brings about two main changes. First, the decline of the insurance benefits is abolished. Under the old system, called Allocation Unique Degressive (AUD), the size of the payments depends on the elapsed duration of unemployment and declines stepwise at the end of predefined intervals. Under the new system, benefits remain at a fixed level for the whole payment period. Second, the new system introduces a range of active labor market policy measures. This includes compulsory meetings on a regular basis with a caseworker. At the first meeting, a personal plan called Plan d'Action Personalisé (PAP) is established. This captures in a contract the details about the degree of assistance provided by the caseworker to the unemployed as well as the targeted job type and the region of search. This contract is updated periodically if the individual remains unemployed, typically every six months. During the first meeting the unemployed is also assigned to one of different types of services such as counseling and training, see Freyssinet (2002) for a detailed description of the reform.

The PARE reform has two distinguishing features. First, individuals whose unemployment spells started before the implementation of the reform and were still unemployed during its commencement were given the option to choose whether they want to stay in the AUD regime or switch to PARE. If an unemployed decides to stay in AUD, his benefits payments continue to follow the decline scheme and no further changes of the status quo take place. If an unemployed individual decides to switch to PARE, his benefit payments are fixed at the latest level paid and no further decline occurs until the end of the payment period (or unemployment exit). The individuals indicate their decision per mail. The choice option does not apply to spells starting after the 1st of July 2001, the day of coming into force of PARE. All new unemployed are automatically assigned to the new system.

The second distinguishing feature of the new system is that although the meetings with the case worker were mandatory, there was no actual monitoring of the job search efforts. Furthermore, the individuals could generally refuse to take part in assigned training or counseling measures without incurring any sanctions. Thus, a relative attractiveness of the benefits level in PARE is not counterbalanced by an increase in monitoring.

Ex ante it is not clear what the overall effect of the reform is. On the one hand, abolishing the benefits decline removes an incentive for a high search effort. Therefore it can be expected that the exit rate from unemployment to employment will decrease. This intuition is incorporated in theoretical models on optimal unemployment insurance design, see e. g. Pavoni and Violante (2007). There is also some empirical evidence for this in the French context, for example in Prieto (2000) and Dormont et al. (2001). These papers use a parametric specification (of PH and MPH models, respectively) to compare the French unemployment insurance system from 1986-1992, which is characterized by a single drop in benefits, with its successor, the AUD system with stepwise declines. However, Le Barbanchon (2012) finds no effect of a prolonged potential duration of the unemployment payments on exit out of unemployment. He uses[15] a regression discontinuity design, with past employment duration as a forcing variable.

On the other hand, increased usage of active labor market policy measures is supposed to increase the exit rate to employment. A vast body of empirical literature investigates the effects of training, counseling and subsidized wages on the employment dynamics; see e.g. Heckman et al. (1999) and Kluve (2010) for overviews. Job search assistance often has a small to modest positive impact on the exit rate to work. Crépon et al. (2005) find a significantly positive impact of counseling on the exit rate to unemployment in France.

3.2. **The data.** The data sample we use is taken from a matching of two administrative data sets: the Fichier Historique (FH) data set, 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) data set, 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 socio-economic status of the individuals ,namely age, gender, marital status, number of children, educational level, professional experience, description of the job position/type in the last employment spell, reason for entering unemployment, exit direction (out of unemployment), and unemployment history. Details about the construction and content of the variables are in Appendix A.2.

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 3.4.

---

[15]The data set we use in this paper is a subset of the data used in Le Barbanchon (2012).

The choice of the cohorts is restricted by the available data. There is no administrative variable that captures the compliance status of the unemployed. Moreover, for budgetary reasons and due to capacity constraints, there was time variation in the elapsed duration at which the caseworker interview and other job search assistance measures took place for those exposed to the new regime. Thus, some individuals might have exited the state of unemployment before those events. 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 AUD coincides with the start of the reform exposure. This enables us to observe the compliance status.[16] Its inflow is six months before the start of PARE.[17] The choice of the comparison cohort (the untreated) is more flexible as we do not need to observe the compliance. The main concerns are macroeconomic: a good choice of a cohort does not violate the randomization assumption. Business cycles or mass layoffs due to bankruptcies of large firms are examples for possible causes for structural changes in the distribution of heterogeneity in the unemployment 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 3.4 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.

3.3. **Estimation results.** We now turn to our main results. With the choice of cohorts described in section 3.2, the treatment effect which we estimate is equal to

$$(3.1) \qquad TE(6, a) = P(T(6) \in [6, 6 + a] \mid T(6) \geq 6) - P(T(9) \in [6, 6 + a] \mid T(9) \geq 6),$$

where a varies between 0 and 3 months. The upper limit three months follows from the time difference of the inflows of the two cohorts. Any comparison beyond this interval would involve two treated groups. The treatment effect gives thus the difference in the probabilities to find a job in the interval $[6, 6 + a]$, conditionally on surviving up to the 6th month, between the old and the new system. Letting a go to 0 gives the average causal effect on the hazard rate.

The result is shown on figure 1. On the x-axis time is measured in days. Each (x, y)-point represents a pair $(a, ATE(a))$. The estimated treatment effect is positive and increasing

---

[16]One may also consider subsequent elapsed durations at which declines take place, but this would be at the cost of having fewer observations.

[17]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 instead does not affect the results. We exclude elderly unemployed for whom the time length differs from 6 months.
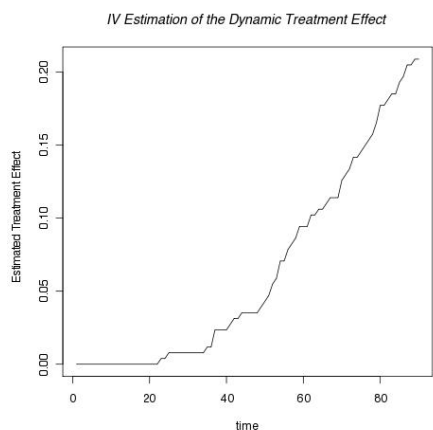
FIGURE 1. An IV estimator of the treatment effect. Time measured in days.

which indicates that the program was effective. The results are similar for different subpopulations, see figures 2a and 2b. Overall, our results are compatible with the findings
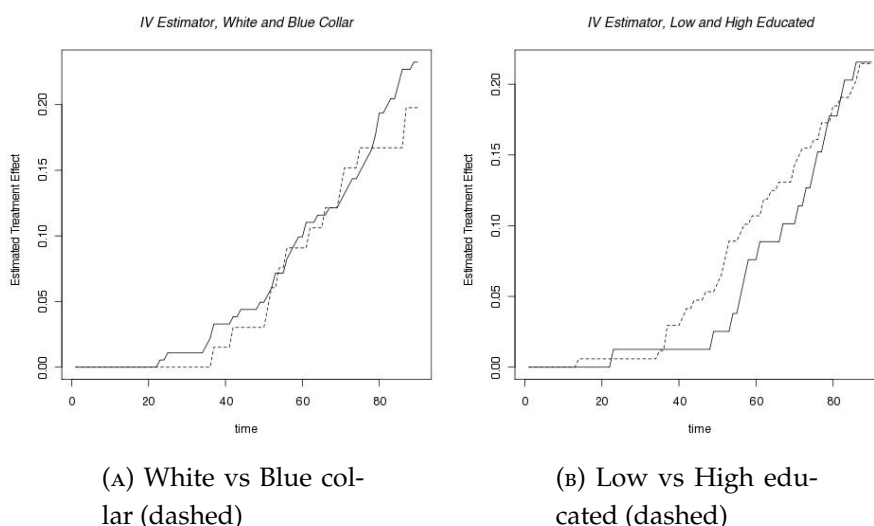

(A) White vs Blue collar (dashed)

(B) Low vs High educated (dashed)

FIGURE 2. Estimates conditional on qualification and education

in the existing literature on the effect of the PARE reform. Fougère et al. (2010) find a positive and significant effect of the reform on the return to employment. They specify the assignment of individuals to treatment types (conditional on being in the new reform) as a discrete choice, and the duration of unemployment as a MPH model. In their analysis, they use spells beginning before and after the day of the reform. Debauche and Jugnot (2007) does not find any evidence that the PARE program accelerates the return to employment. Similarly to Fougère et al. (2010), they use both spells beginning before and after the begin of the reform. For their analysis, they use a nonparametric Kaplan-Meier estimator and simulated Cox-Model, and point out that the effect of the reform on the unemployment hazard cannot be clearly separated from changes in the macroeconomic conditions. Further, Crépon et al. (2012) find that training does not accelerate exit out of unemployment but increases the length of the subsequent employment spell.

They utilise the (semi-) parametric framework of the timing-of-event method. They use all spells with inflow between 2001 and 2005. The authors' finding is related to the idea that training increases human capital and improves matching process between firms and unemployed. Next, Crépon et al. (2005) find that three out four counselling schemes have a positive effect on the unemployment hazard.[18] One possibility is that jointly these two types of services offset the negative impact of the generous benefit system. Another possibility is that the financial side of the reform has little or no stimulus on the behaviour of the individuals. Evidence for this possibility is provided by the study of the French unemployment system in Le Barbanchon (2012). He finds no negative effect of the longer potential duration of benefits on the exit out of unemployment using data for the period 2000-2002.

Remark: it must be noted that our evaluation subsumes several different treatments into one single treatment. We interpret our evaluation as averaging over the different treatment schemes. Furthermore, even if there was no real monitoring, it is likely that the regular meetings with the caseworker were perceived as monitoring. Averaging over different treatment types is not uncommon in the literature, see for example Blundell et al. (2004) and Van den Berg et al. (2014) for evaluations of the New Deal for Young People program in the UK.

3.4. **The validity of the assumptions.** We start with the randomization assumption A3. Since $T(s), S(t)$ and V are in general unobserved, A3 cannot be tested directly. However, we may verify that the cohorts are similar at their inflows in terms of

  (1) the distributions of the individual observed characteristics,
  (2) the layoff reasons and
  (3) the macroeconomic conditions

These comparisons provide indirect evidence for A3. We first perform a chi-square test for equality of distributions of level of education, years of experience, number of children and gender. The corresponding p-values are 0.6037, 0.98, 0.5112 and 0.581, which indicates that the differences between these distributions are statistically insignificant. This is reflected in their histograms, see figures 3a - 6b. 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 7a-7b. Third, the next we show 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[19].

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. Unemployed were

---

[18]Similarly to the other related papers, this study estimates a MPH model for the hazard of unemployment. Only spells that begin after the reform are taken into consideration.
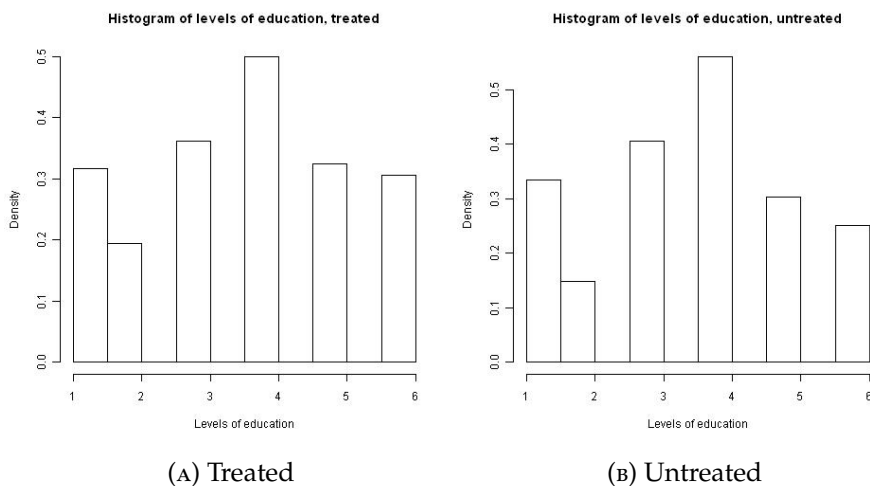
[19]Source: http://www.insee.fr/en/bases-de-donnees/bsweb

(A) Treated   (B) Untreated

FIGURE 3. Histograms level of education
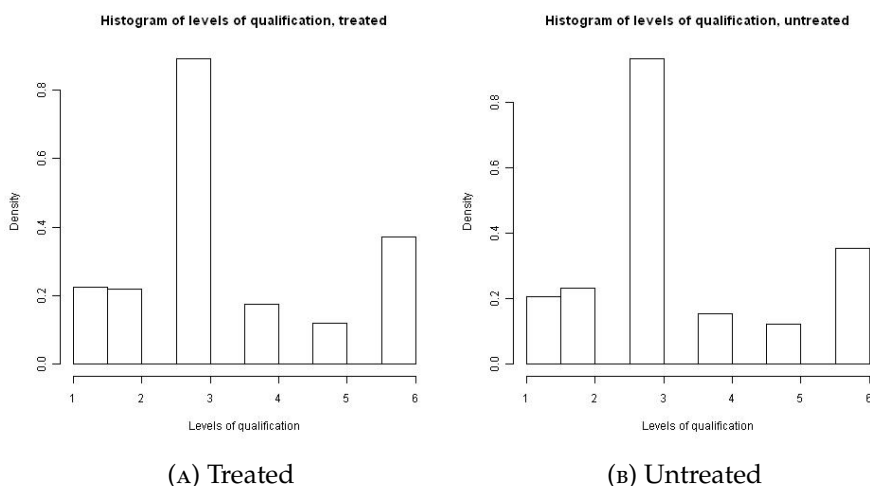

(A) Treated   (B) Untreated

FIGURE 4. Histograms level of qualification

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.

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. In subsection 3.6, we conduct a simulation study, in which plausible deviations from the independence assumptions are generated. It turns out that the estimator is robust towards such violations.
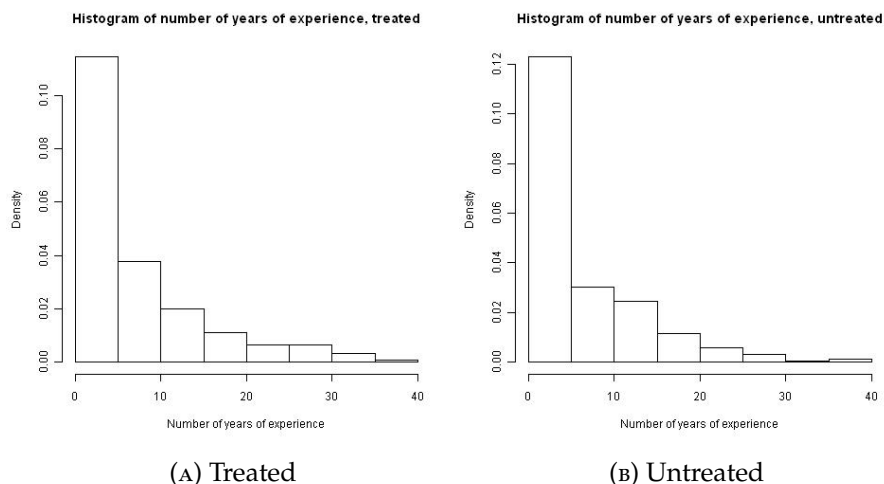
(A) Treated   (B) Untreated

FIGURE 5. Histograms of years of experience

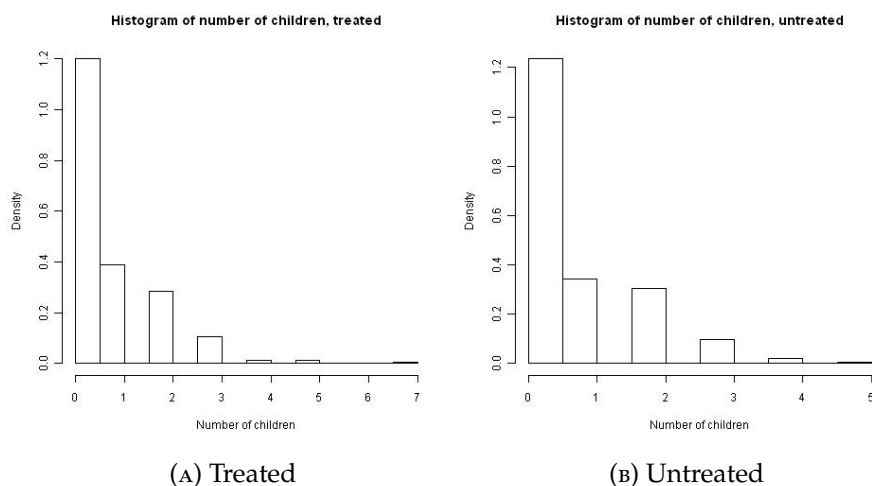(A) Treated   (B) Untreated

FIGURE 6. Histograms of number of children

Remark: one implicit implication of the assumptions A2 and A4 is that noncompliers who refuse the treatment behave in the same way as if they were not assigned to the treatment. It is plausible, particularly in the cases in which individuals do not comply **in order** not to change their behavior. In the PARE reform, it is plausible to assume that individuals do not comply because they anticipate a fast exit and because they want to avoid higher search effort or other related participation costs; see the next subsection for an analysis of noncompliance. In both cases, assignment to the treatment together with a selection out of it is not likely to change their behavior.

In sum, the assumptions adopted for identification and estimation of the treatment effect of the PARE reform can be considered as plausible.

3.5. **Analysis of endogeneity.** In this subsection we tackle the static endogeneity issue arising from noncompliance. Noncompliance is important not only for the evaluation of a program but also in the light of its effectiveness. The non take-up of a policy often

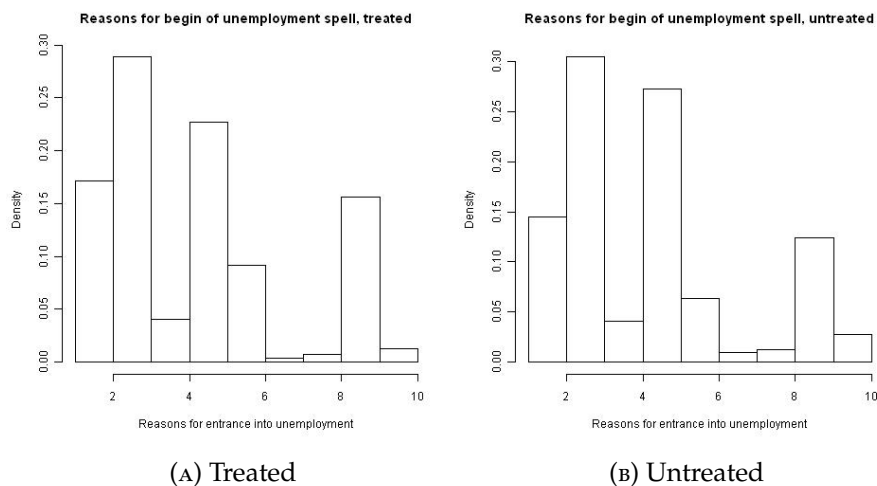(A) Treated   (B) Untreated

FIGURE 7. Histograms of layoff reasons

reduces the effectiveness of a program, see e. g. Blasco (2009). It is therefore important to understand what drives noncompliance.

We start with an estimation of the naive treatment effect (2.8). The corresponding estimator is defined as

$$(3.2) \qquad \widehat{NE}(t,a) := \widehat{P}(T \in [t, t+a] \mid S = t, Z = t) - \widehat{P}(T \in [t, t+a] \mid S = \infty, Z = t),$$

where t is equal to 6 months and a varies between 1 day and 3 months and the separate probabilities are estimated with a Kaplan-Meier estimator. (3.2) amounts to a direct comparison of the average outcome of compliers and noncompliers from the cohort $\{Z = 6\}$. The estimate is shown in figure 8. It is positive and increasing until the 80th day after
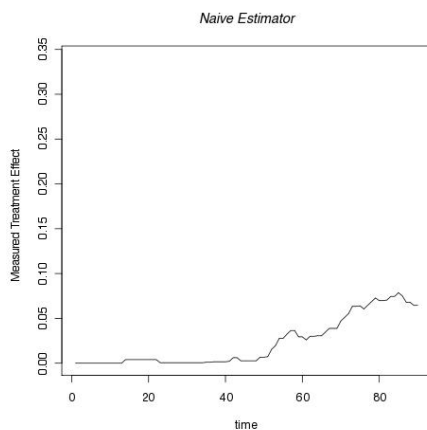

FIGURE 8. A naive estimator: noncompliers as control group. Time measured in days.

treatment (which is the 260th day of unemployment), and then slightly decreasing. At the first 40 days after treatment the effect is practically zero, at its maximum it is around 0.08, and at day 60 after treatment (that is, after 8 months of unemployment) around 0.025. This implies that the probability for a complier to find a job before the end of the first month

after treatment, conditional on having been unemployed for 6 months, is almost the same as for a noncomplier, before the end of the second month it is with 2.5 percentage points higher, and at its peak it is 8 pp. higher.[20] As a result, if we evaluate PARE using the naive estimator (3.2), we would conclude that the reform was beneficial for the duration of unemployment but that the effect is rather modest.

Using the methods developed in section 2.5, we answer now the following questions:

- Is the non take-up of PARE driven by an endogenous selection?
- If yes, in which direction is the bias of the naive estimator caused by this endogenous selection?

To answer the first question, we perform the simplified test for exogeneity from section 2.5. The null hypothesis is

(3.3) $$H_0: \qquad \bar{F}_1(6 \mid S = \infty) - \bar{F}_2(6).$$

It amounts to comparing the survival function at $t = 6$ of the noncompliers from cohort $\{Z = 6\}$ with the survival function at $t = 6$ of the whole cohort $\{Z = 9\}$. The test statistics is defined as $\mathcal{T} = \widehat{\bar{F}}_1(T > 6 \mid S = \infty) - \widehat{\bar{F}}_2(6)$, where $\widehat{\bar{F}}_i, i = 1, 2$ are the Kaplan-Meier estimators of $\bar{F}_i$.

The test rejects the null at 5 % level. As a result, untreated compliers and noncompliers are significantly different in terms of potential outcomes, which induces a static selection bias in the naive estimator. To evaluate the bias of (3.2), we plot $\widehat{TE}(6, a)$ and $\widehat{NE}(6, a)$ for a varying between 0 and 3 months. The result is shown in figure 9. $\widehat{NE}(6, a)$ has a
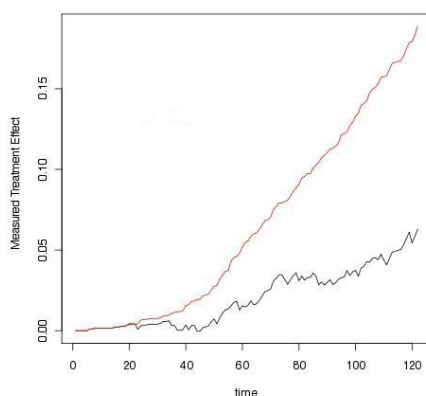

FIGURE 9. Comparison of the IV estimator (red line) and the naive estimator (black line). Time measured in days

negative bias for all a. To find the reason for this negative bias, it is helpful to interpret it in the frame of existing studies on policy take-up, see e.g. Moffit (1983), Currie (2004) and Blasco (2009). An empirical analysis of the take-up of the PARE reform is done by Blasco (2009), who also uses a theoretical model. She finds stigma, informational issues and the

---

[20]Controlling for observed covariates such as gender, education and type of job (white vs. blue collar) yields similar results, see figures 14, 16a and 16b in appendix A.3. Due to the small sample size available, we conduct the study mainly unconditionally.

expectation of a soon exit to be the main reasons for noncompliance. One explanation for the negative bias in line with Blasco (2009) would be therefore the expectation of a short spell among noncompliers. Individuals who anticipate to find quickly a job or who have even already signed a contract at the time of the reform start would be reluctant to comply since they wouldn't benefit from the generosity of the new program. Thus, there is a selection of quick exits into the group of noncompliers which leads to the negative bias of the naive estimator.

3.6. **Dependent Censoring: a Simulation Study.** To assess the impact of the assumption of independent censoring, a small simulation study is conducted. 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 3.5 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 3.3[21]. 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 10 shows the theoretical treatment effect,
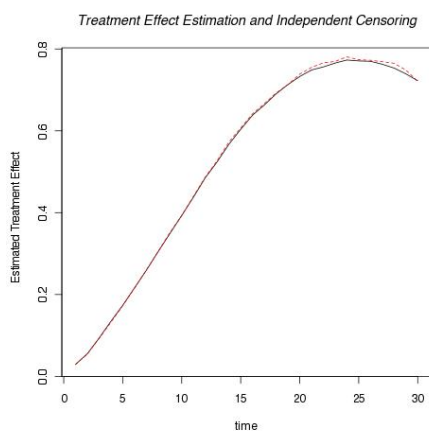

Treatment Effect Estimation and Independent Censoring

FIGURE 10. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20).

---

[21]Negative values are replaced by their absolute values.

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.

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 11. The solid black line is theoretical effect. The figure reveals the relationship between bias of the treatment effect and dependence
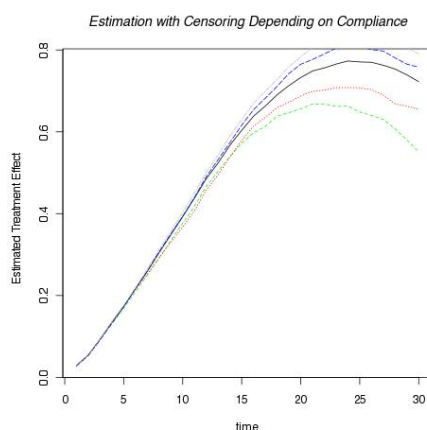

FIGURE 11. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical treatment effect. Different curves correspond to different dependences of censoring and compliance, see table 1. The solid black line 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 distribution $T \leq 40$	Censoring distribution $T > 40$
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 12. 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).
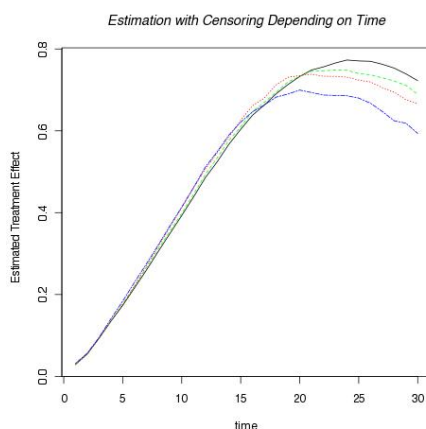

FIGURE 12. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20).The black solid line is the theoretical treatment effect. Different curves correspond to different dependences of censoring and time, see table 2. The solid black line is theoretical effect.

It is interesting to simulate and analyze a combination of these two types dependences. We simulate four patterns of such an interplay. The concrete distributions are described in table 3. The results are shown in figure 13. The blue and the grey lines are closer to the theoretical effect than the other two estimators. This indicates, that a violation 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

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, $T \leq 30$
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.

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.
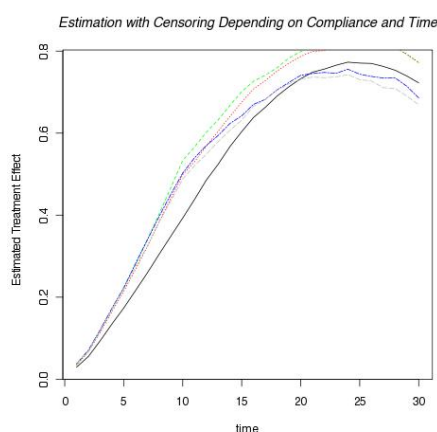

FIGURE 13. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical effect in the absence of censoring. Different curves correspond to different dependences of censoring and time, see table 2. The solid black line is theoretical effect.

4. SUMMARY AND DISCUSSION

In this paper we developed a nonparametric IV framework for the evaluation of dynamic treatment effects. Our methods solve the problems of dynamic and static endogeneity and allow for censoring. The corresponding estimators have a natural interpretation and are related to the Wald-type statistics. We also suggest a framework for analysis of noncompliance. We used our methods to evaluate the French labor market reform PARE. The estimated effect of the reform on the conditional survival function of the unemployment variable is positive, which is in line with the findings in the existing literature. In an exhaustive study, we showed that the assumptions for our approach are valid. Our results reveal that neglecting of endogeneity would lead to a negative bias. An interesting question for future research would be to incorporate equilibrium effects. Comprehensive

policy reforms are likely to induce equilibrium effects through positive or negative externalities. It is often desirable to distinguish between the direct effects of a reform and the equilibrium effects. More work on this topic has to be done.

A.1. **Proofs of propositions.** **Proof of Proposition 2.1**

First we show that from the no anticipation assumption the following result holds:

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

This is so because

$$P(T(t) \geq t \mid X, S(t) = t, V) = \exp(-\Theta_{T(t)}(t \mid X, S(t) = t, V))$$

$$\overset{\text{No anticipation}}{=} \exp(-\Theta_{T(t')}(t \mid X, S(t) = t, V)) = P(T(t') \geq t \mid X, S(t) = t, V),$$

so that we obtain

$$P(T(t) \geq t \mid X, S(t) = t) = \mathbb{E}\left[I_{\{T(t)\geq t\}} \mid X, S(t) = t\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[I_{\{T(t)\geq t\}} \mid X, S(t) = t, V\right] \mid X, S(t) = t\right]$$

$$= \mathbb{E}\left[P(T(t) \geq t \mid X, S(t) = t, V) \mid X, S(t) = t\right]$$

$$= \mathbb{E}\left[P(T(t') \geq t \mid X, S(t) = t, V) \mid X, S(t) = t\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[I_{\{T(t')\geq t\}} \mid X, S(t) = t, V\right] \mid X, S(t) = t\right] = P(T(t') \geq t \mid X, S(t) = t)$$

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 (A.1), we show $F_{V\mid T(t)\geq t, X, S(t)=t} = F_{V\mid T(t')\geq t, X, S(t)=t}$. Let B be a Borel set. With result (A.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 $F_{V\mid T(t)\geq t, X, S(t)=t} = F_{V\mid T\geq t, X, S=t, Z=t}$. 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

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

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

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

□

Proof of corollary 2.1

With proposition 2.1,

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

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

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

Lemma A.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*

(A.2)

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

(A.3)

$$P(T(t') \in B \mid T(t') \geq t, X, S(t) = \infty) = P(T \in B \mid T \geq t, X, S = \infty, Z = t) \qquad and$$

(A.4)

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

Proof

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

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

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

$$= P(T(t') \in B \mid X, S(t) = \infty),$$

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

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

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

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

Proof

First, it holds

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

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

where the second equality follows with assumptions A1-A4.

Next,

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

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

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

where the second equality holds due to no anticipation.

Proof of proposition 2.2

First, write

(A.7) $\qquad P(T(t') \in [t, t + a) \mid T(t') \geq t, X)$

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

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

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 (A.7). Plugging in the results of lemma A.1 and lemma A.2, we obtain for $F_{C,0} := P(T(t') \in B \mid T(t') \geq t, X, S(t) = t)$

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

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

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 2.7

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 2.7 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

$$(A.8) \qquad b^{-2}(\Psi^* - \Psi) = \frac{b^{-2}}{\widehat{p}_1}((\theta_1^* - \theta_1) - (\theta_2^* - \theta_2)) + b^{-2}(\theta_1 - \theta_2)(\frac{1}{\widehat{p}_1} - \frac{1}{p_1}).$$

The second term on the right-hand side of (A.8) is equal to $o_p(1)$ when b is of order $O(n^{-1/(q+5)})$ or $o(n^{-1/(q+5)})$. Proposition 2.7 ii) follows with Theorem 1 b) in Nielsen and Linton (1995). Finally, proposition 2.7 iii) follows directly from the adapted proof of Theorem 1 c) Nielsen and Linton (1995) and the continuous mapping theorem.

A.2. **Description of variables.** The variables used in our 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 technicier supérieur and diplôme univeritaire 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 Crépon et al. (2010). 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.
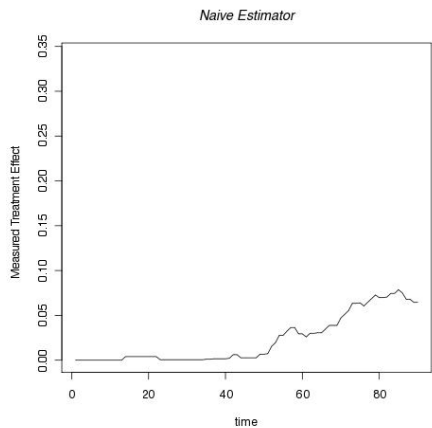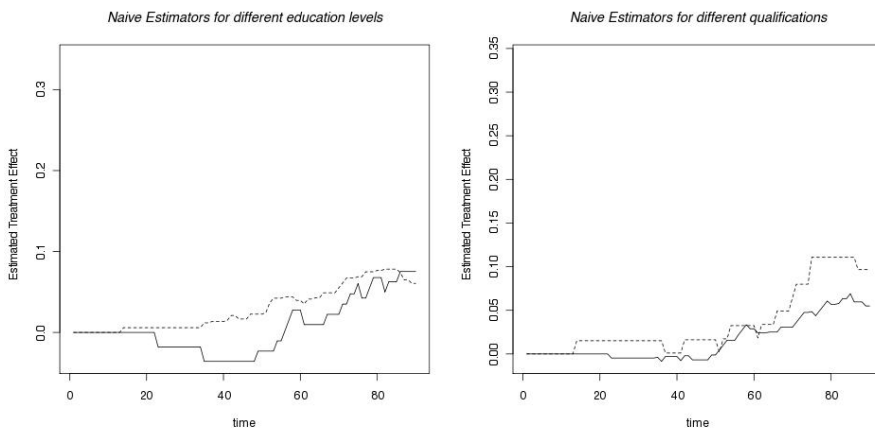
A.3. **Analysis of endogeneity.**

FIGURE 14. A naive estimator. Male vs. Female. Time measured in days.

FIGURE 15. A naive estimator for subgroups

(A) Education. Low educated dashed line

(B) Qualification. Blue collars dashed line

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.

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., Costa Dias, M., Meghir, C., and Van Reenen, J. (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, 2(4):569–606.

Chesher, A. (2002). Semiparametric identification in duration models. CeMMAP working paper CWP20/02, Centre for Microdata Methods and Practice, London, UK.

Cox, D. (1962). *Renewal Theory*. Methuen, 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. (2010). Analyzing the anticipation of treatments using data on notification dates. IZA Discussion paper 5265, Institute for the Study of Labor (IZA), Bonn, Germany.

Currie, J. (2004). The take-up of social benefits. Technical report, Conférence an l'honneur de Eugène Smolensky.

Debauche, E. and Jugnot, S. (2007). Les effets du projet d'action personalisé sur les sorties des listes de l'ANPE. Working paper, DARES, Paris, France.

Dormont, B., Fougère, D., and Prieto, A. (2001). L'effet de l'allocation unique dégressive sur la reprise d'emploi. *Économie et Statistique*, 343(1):3–28.

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.

Fougère, D., Kamionka, T., and Prieto, A. (2010). L'efficacité des mesures d'accompagnement sur le retour à l'emploi. *Review Économique*, 61(3):599–612.

Freyssinet, J. (2002). La réform 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.

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.

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.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6 of *Handbook of Econometrics*, chapter 70. Elsevier.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

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.

Kluve, J. (2010). The effectiveness of European active labor market programs. *Labour Economics*, 17:904–918.

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.

Moffit, R. (1983). An economic model of welfare stigma. *The American Economic Review*, 73(5):1023–1035.

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.

Pavoni, N. and Violante, G. L. (2007). Optimal welfare-to-work programs. *Review of Economic Studies*, 74(1):283–318.

Pearl, J. (2000). *Causality. Models, reasoning, and inference*. Cambridge University Press.

Prieto, A. (2000). L'impact de la dégressivité des allocations chômage sur le taux de reprise d'emploi. *Revue Économique*, 51(3):523–534.

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.

Tchetgen, E. J., Walter, S., Martinussen, T., and Glymour, M. (2014). Instrumental variable estimation in a survival context. Biostatistics working paper series 179, Harvard.

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*.