

VU Research Portal

Nonparametric instrumental variable methods for dynamic treatment evaluation

Van Den Berg, Gerard J.; Bonev, Petyo; Mammen, Enno

published in

Review of Economics and Statistics
2020

DOI (link to publisher)

[10.1162/rest_a_00843](https://doi.org/10.1162/rest_a_00843)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Van Den Berg, G. J., Bonev, P., & Mammen, E. (2020). Nonparametric instrumental variable methods for dynamic treatment evaluation. *Review of Economics and Statistics*, 102(2), 355-367.
https://doi.org/10.1162/rest_a_00843

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

NONPARAMETRIC INSTRUMENTAL VARIABLE METHODS FOR DYNAMIC TREATMENT EVALUATION

Gerard J. van den Berg, Petyo Bonev, and Enno Mammen*

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

I. Introduction

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

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

groups in a dynamic setup. Finally, when the outcome of interest is a duration variable, identification might be hampered by right-censoring.

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

At the core of our method is a dynamic potential outcomes framework. A treatment is assigned at some random elapsed duration of unemployment. The interest is in the effect of this treatment on posttreatment outcomes such as posttreatment unemployment duration. A major question in this setup is how to define a meaningful treatment effect. While the standard static literature defines those who are not observed to enroll in the treatment as nontreated, in a dynamic setting, this approach leads to conditioning on future successful outcomes (Sianesi, 2004).¹ To avoid this problem, we follow much of the literature on dynamic treatment effects and focus on treatment effects on the subgroup of individuals who remain unemployed at least until the treatment (Sianesi, 2004; Vikström, 2017; van den Berg, Bozio, & Costa Dias, 2020).

Our main contribution is to extend the dynamic treatment evaluation framework to allow for endogenous noncompliance. Noncompliance has been largely ignored by the literature on dynamic treatment effects, despite the attention it has received in the static literature. We develop a dynamic concept of noncompliance that allows the individual to change preferences in an arbitrary way over time. The intuition is that whether an unemployed accepts to participate in a labor market program depends on her subjective probability of finding a job, and this probability might change with the elapsed duration of unemployment.

Our method relies on two major assumptions. The first one is exogenous variation in the timing of treatment assignment. As a motivation, consider a case worker who is responsible for a large pool of unemployed. Then, conditional on characteristics of the unemployed, the case worker might assign the order in which the unemployed are advised idiosyncratically (Sianesi, 2004). Our strategy is also motivated by so-called phased-in experiments, in which randomly selected late recipients provide a control group for randomly selected early recipients (Duflo, Glennerster, & Kremer, 2007). Finally, when a policy is administered at a single point in time, then the presence of cohorts who enter unemployment at

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

Received for publication July 5, 2016. Revision accepted for publication January 11, 2019. Editor: Bryan S. Graham.

*van den Berg: University of Bristol, University of Groningen, IFAU-Uppsala, IZA, ZEW, and J-PAL; Bonev: University of St. Gallen; Mammen: Heidelberg University.

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

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00843.

different times might also give rise to quasi-experimental variation in the time to treatment.

The second one is a dynamic consistency assumption commonly referred to as “no anticipation” in the literature (Abbring & van den Berg, 2003). This assumption requires that if two labor market treatments coincide up to a duration t , then the hazards of unemployment duration under the two treatments should be the same for each pretreatment duration. With forward-looking individuals, this may require that individuals do not anticipate the (time of) treatment or do not act on their knowledge. The no-anticipation assumption has been used throughout the literature on dynamic treatment effects (Crépon et al., 2009; Crépon et al., 2018; Vikström, 2017). Moreover, it is an implicit assumption in some standard static approaches, such as the difference-in-difference (DiD) and the synthetic control approaches. It is also an implicit assumption in the phased-in experiments, where ignoring the time dimension effectively subsumes the no-anticipation assumption into the randomization assumption. Our paper puts phased-in experiments into a dynamic framework and makes the link to no anticipation explicit.

Our identification strategy consists of two steps: a dynamic and a static one. In the dynamic step, initial randomization of the assignment and no anticipation ensure that dynamic selection follows the same pattern for early and late treatment recipients at each pretreatment duration.² This allows for a comparison of treated and not-yet-treated individuals at the same elapsed duration of unemployment. In a second step, the assignment to treatment is used as an instrument for the endogenous enrollment into treatment. Information on observed compliance from the early recipients is used to identify the outcome distribution of late, not-yet-treated recipients.

With these two steps, our paper provides a link between the dynamic treatment evaluation literature (Vikström, 2017; Sianesi, 2004; Lechner, Miquel, & Wunsch, 2011; van den Berg et al., 2020) and the standard static local average treatment effects (LATE) literature (Imbens & Angrist, 1994; Imbens & Rubin, 1997). Identification is local in the sense that at each point in time the effect is identified only for those who would comply at this particular time. The static notion of location is thus extended with a time component. The corresponding estimators can be viewed as dynamic Wald estimators. Moreover, in a setup where time to treatment in the control group approaches the time to treatment in the treatment group, our method can be extended to a dynamic fuzzy regression discontinuity (RD) approach.

Our paper contributes to the literature on IV in survival analysis (Eberwein, Ham, & LaLonde, 1997; Robins & Tsiatis, 1991; Chesher, 2002; Bijwaard & Ridder, 2005; Bijwaard, 2009; Tchetgen et al., 2014). Much of this literature is surveyed by Abbring & van den Berg (2005). Typically these studies adopt a semiparametric or a paramet-

ric model structure. Our model, on the contrary, is fully nonparametric. Dynamic discrete choice models also deal with identification of dynamic treatment effects (Heckman & Navarro, 2007). Contrary to our approach, those papers rely on period-specific exclusion restrictions, as well as on restrictive separability and identification-at-infinity assumptions. Our approach is also related to the literature on duration models with a mixed proportionate hazard (MPH) structure and time-varying covariates—in particular, to the important paper of Hausman and Woutersen (2014). A thorough overview of this literature is provided in Hausman and Woutersen (2008). Similar to our approach, identification in Hausman and Woutersen (2014) relies on variation in the time to treatment. The major difference is how dynamic selection is handled. While we handle dynamic selection by assuming no anticipation of the treatment, dynamic selection is modeled explicitly in Hausman and Woutersen (2014) through the semiparametric assumption on the hazard and in particular through a separability assumption on the unobserved heterogeneity. In addition, their rank estimator utilizes the monotonicity of the hazard function in the observed covariates X implied by a parametric assumption. Another paper that uses variation in the time to treatment as a source of identification is Abbring and van den Berg (2003). Contrary to our nonparametric model, both the treatment process and the effect of the treatment on the outcome duration are modeled within the semiparametric MPH framework. Finally, our dynamic RD extension is related to the static RD approach in Hahn, Todd, and Van der Klaauw (2001).

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

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

II. Econometric Framework

A. Treatment Effects

For illustrative purposes, we build our exposition on a labor market example. Suppose we observe a sample of n individuals who are searching for a job. As part of active labor

²van den Berg et al. (2020) derive this result in a context with full compliance. We generalize their result to a setup with endogenous noncompliance.

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

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

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

$$T_i(s, z) = T_i(s) \quad \text{for each } s, z \in \mathbb{R}^+ \cup \{\infty\}. \quad (1)$$

Equation (1) prevents the assignment to treatment from directly affecting the outcome: Z_i can influence the duration only through the actual pretreatment duration S_i . This assumption is plausible in our setting since we restrict Z_i and S_i to realize simultaneously from the viewpoint of the unemployed. This is in contrast to the case in which the assignment to treatment realizes prior to the treatment.⁴

The second aspect that differs from the static framework is that a simple comparison of treated and nontreated within the treatment definition window possibly leads to a bias.⁵ In particular, if an individual finds a job prior to the treatment, S_i will be censored by T_i and therefore unobserved. Considering these individuals as nontreated effectively conditions on their future successful outcomes (Sianesi, 2004). We therefore follow the approach chosen by much of the literature on dynamic

treatment effects and condition on survival in unemployment up to treatment (Vikström, 2017; Sianesi, 2004).⁶

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

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

$$\begin{aligned} TE(t, t', a) &= \mathbb{E}[P\{T(t) \in [t, t+a] \mid T(t) \geq t, X, V, S(t) = t\} \\ &\quad - P\{T(t') \in [t, t+a] \mid T(t') \geq t, X, V, S(t) = t\} \\ &\quad \mid T(t) \geq t, X, S(t) = t], \end{aligned} \quad (2)$$

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

The interpretation of equation (2) is as follows. Since $t + a < t'$, all individuals assigned to be treated at t' are not yet treated in $[t, t+a)$. Thus, $P\{T(t') \in [t, t+a] \mid T(t') \geq t, X, V, S(t) = t\}$ can be interpreted as the individual potential outcome of a nontreated individual who would remain unemployed under the treatment t' at least until t . $P\{T(t) \in [t, t+a] \mid T(t) \geq t, X, V, S(t) = t\}$ is the corresponding outcome of a treated at t individual. Thus,

$$\begin{aligned} &P\{T(t) \in [t, t+a] \mid T(t) \geq t, X, V, S(t) = t\} \\ &\quad - P\{T(t') \in [t, t+a] \mid T(t') \geq t, X, V, S(t) = t\} \end{aligned} \quad (3)$$

is the additive individual treatment effect on the probability of leaving unemployment in $[t, t+a)$ for an individual who is still unemployed at t . Expression (2) is an average of equation (3). The average is built with respect to the distribution of unobserved heterogeneity V among the treated survivors, $F_{V|T(t) \geq t, X, S(t) = t}$.⁸ Proposition 1 in section III shows that under

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

⁷We suppress the index i for notational simplicity.

⁸Note that equation (2) can be also written as

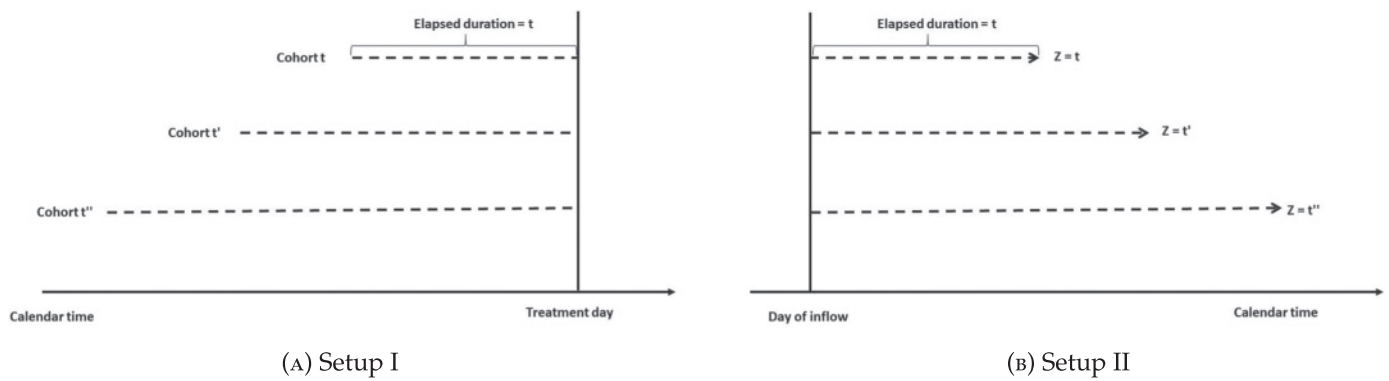
$$\begin{aligned} &\mathbb{E}_{V|T(t) \geq t, X, S(t) = t}[P\{T(t) \in [t, t+a] \mid T(t) \geq t, X, V, S(t) = t\} \\ &\quad - P\{T(t') \in [t, t+a] \mid T(t') \geq t, X, V, S(t) = t\}]. \end{aligned}$$

³See Sianesi (2004), Crepon et al. (2018), and Carney and Ramia (2011).

⁴The standard example of nonsimultaneous realizations is when the individual is warned about a future sanction due to noncompliance with UI rules (see Crepon et al., 2018). In this case, Z_i might have a direct effect on the outcome, which is often referred to as a threat effect.

⁵The treatment definition window is the period of time used in a static setup to define treatment status.

FIGURE 1.—TWO EMPIRICAL SETUPS



Setup I: Many cohorts, common point in time of treatment. Setup II: Simultaneous inflow, variation in time to treatment.

certain assumptions, this is also the distribution of V among all survivors at t .

Conditioning on $S(t) = t$ restricts the evaluation of the effect on the t -compliers—the individuals who would take the treatment at an elapsed duration t if it was assigned to them. Offering the treatment only once means that we allow for only one-sided noncompliance. With one-sided noncompliance, the set of treated compliers coincides with the full set of treated (at elapsed duration t). Therefore, equation (2) can be interpreted as a treatment effect on the treated.

By conditioning on survival up to treatment and focusing on the posttreatment outcome, equation (2) closely resembles the definition of a treatment effect in the dynamic impact evaluation literature. The novel aspect that our paper introduces is to explicitly allow for noncompliance. The stochastic process $\{S(t)\}_{t \geq 0}$ is a generalization of the static compliance model in the LATE literature (Imbens & Angrist, 1994). Equation (2) thus provides a natural link between the dynamic treatment evaluation literature and the static LATE.

Remark 1. It is clear from definition (2) that the duration outcome $T(s)$ can be replaced by an arbitrary posttreatment outcome $Y(s)$. In fact, our identification strategy presented in the next section does not rely on the outcome being a duration variable.⁹

Remark 2. For any two t, t' , there is a variety of possible treatment effects, one for each $a \in (0, t' - t)$. In addition, an evaluation of the total effect of a policy might involve averages over all t and t' .

Remark 3. An important special case of equation (2) is the limit case $a \rightarrow 0$. It amounts to a treatment effect on the hazard function. We devote special attention to this case in the next section and in the appendix.

Remark 4. The precise interpretation of equation (2) is less straightforward when the treatment is not instantaneous. As an example, long-term training measures induce a lock-in effect. One approach followed in the literature is to treat the lock-in effect as constituent of the total treatment effect (see

Sianesi, 2004). A second problem, however, is that the length of a noninstantaneous treatment might also matter. We therefore concentrate on instantaneous treatments, such as short-term training, counseling, and other activating measures. In the context of the labor market example, such focus is not an important restriction, as there is a general trend in labor market policies toward short-term activation and reemployment ALMPs (Biewen et al., 2014).

Remark 5. Expression (2) implies that we treat $T(s)$ as a random variable even when we condition on all observed and unobserved characteristics X, V . By doing that, we follow the general approach in mixture duration models pioneered by Lancaster (1979). The underlying assumption is that the randomness in $T(s)$ comes from some intrinsic uncertainty in the transition, not observed and controlled by the individual.¹⁰

B. Two Empirical Setups

We focus on two particular empirical setups.

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

⁹Our estimator, however, uses the relation between the outcome and the conditioning set.

¹⁰This distinction, however, is arbitrary and is mainly made for technical reasons (Lancaster, 1990).

Empirical example I: In section VI, we evaluate the French unemployment policy reform PARE, which changed the payment structure of unemployment benefits and introduced ALMPs such as training. The new regime was effective from July 1, 2001 (comprehensive treatment assignment). Non-compliance was possible.

Setup II: Phased-in treatment. Consider a group of individuals who enter the state of interest simultaneously (a single cohort). Treatment is assigned at different elapsed durations for different individuals (see figure 1b).

Empirical example IIa—Time variation in ALMPs: Often there is time variation in administering ALMPs across individuals (Lalive, van Ours, & Zweimüller, 2005; Sianesi, 2004; Crepon et al., 2018; Abbring & van den Berg, 2003). Time variation might be necessary due to budgetary or other administrative reasons. If a case worker is responsible for a large group of unemployed, then meetings and coaching sessions must be assigned at different dates and hence at different elapsed durations of unemployment Z_i .

Empirical example IIb—Phased-in implementation of social programs and field experiments: Consider a social program characterized by a phased-in implementation: some units (regions or individuals) are assigned to be treated earlier and others later. Phased-in implementation might be necessary for similar reasons as in example IIa. Alternatively, a comprehensive program might be preceded by pilot studies introduced at varying times, which also gives rise to the setup described above. As an example, Blundell et al. (2004) utilize the presence of area-based pilot studies of the U.K. policy reform New Deal for Young People (NDYP) to identify its impact on unemployment.¹¹ An important subcategory is the phased-in (or pipeline, or rolled out) experimental design, in which the treatment is assigned at different times and the order is randomized. The seminal deworming study of Miguel and Kremer (2004) is the standard example here. Not-yet-treated individuals are taken as a control group for already treated individuals. Below, we explicitly formalize the assumptions that are typically made in this literature.

III. Identification of Dynamic Treatment Effects

A. Assumptions

Consider the following assumptions.

Assumption 1: Dynamic Noncompliance. For any t , it holds either $S(t) = t$ or $S(t) = +\infty$.

Assumption 1 defines the possible type of noncompliance. Agents are allowed only to choose between being treated at the assigned point in time and being never treated. It thus precludes the type of choices $S(t) = t'$ for some $t' \neq t$ with

$t' < \infty$. However, it allows individuals to change their preferences over time in an arbitrary way. As an example, for $t < t'$, individual i is allowed to choose $S_i(t) = \infty$ and $S_i(t') = t'$. A noncomplier at t might be a complier at t' . In the context of an ALMP, an unemployed worker's decision whether to participate in an offered training at a given time will depend on the worker's subjective probability of the prospects of getting a job without the training. An unsuccessful period of search might increase the readiness to participate as the individual gets more pessimistic.

Assumption 1 mimics the standard static noncompliance model, in which the treatment is offered only once (Heckman, LaLonde, & Smith, 1999). At the same time, the process $\{S(t)\}_{t \in \mathbb{R}^+ \cup \infty}$ extends the static concept of noncompliance by adding a time structure. For a given point in time t , assumption 1 corresponds to a one-sided noncompliance in the static treatment evaluation literature. One-sided noncompliance precludes the existence of always-takers and defiers.¹² As a result, no monotonicity-type assumption (as the one invoked in static LATE model) is needed for identification.

Examples I and IIa (continued): Assumption 1 is natural in the setup of the PARE reform, in which the treatment is administered at one single point in calendar time. Administrative and legislative rules prevent unemployed from enrolling in treatment earlier or later. In the context of the Swedish ALMPs, however, an assignment to treatment (Z) by a case worker serves only as a recommendation. An unemployed is free to enroll in the treatment earlier or later, or never. Thus, $S(t) = t'$ is possible for $t' < t$ and $t' > t$, which violates assumption 1.

Assumption 2: No Anticipation. Let $\Theta_{T(s)}$ be the integrated hazard of $T(s)$. Then, for each real $t' \geq t \geq 0$, it holds

$$\Theta_{T(t')}(t | X, V, S(t), S(t')) = \Theta_{T(\infty)}(t | X, V, S(t), S(t')). \quad (4)$$

Assumption 2 states that present potential outcomes are not influenced by future events. This can be seen more clearly in the following relationship:

$$\begin{aligned} P(T(t') > t | X, V, S(t), S(t')) \\ = P(T(\infty) > t | X, V, S(t), S(t')). \end{aligned} \quad (5)$$

Equality (5) is equivalent to assumption 2. It states that the individual survival probability up to some earlier time t remains the same under any potential future treatment times t', t'' ($t \leq t', t''$). Conditioning on $S(t), S(t')$ implies that assumption 2 is valid for the subgroups of compliers and noncompliers at t and t' . In a setting with forward-looking individuals, this assumption is satisfied when the information structure is invariant to the potential assignment of the

¹¹A further example for pilot studies is the implementation of the Progres welfare reform in Mexico (Todd, 2007). Another example for phased-in social programs is the study of the effect of titling land properties on labor market participation (Field, 2007).

¹²To see this, note that assumption 1 precludes choices $S(\infty) = t$ for $t < \infty$. See Imbens and Angrist (1994) for the definitions of always-takers and never-takers.

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

No anticipation in setup II. Assumption 2 is implicitly assumed in all experiments with a phased-in design (example IIb). A unit treated at t' can be used at t , $t < t'$, as a control for a unit treated at t only if the not-yet-treated unit does not anticipate the treatment at t' . Violation of this assumption is considered to be one major potential flaw in the evaluation of phased-in treatments (Duflo et al., 2007). Note that phased-in experiments are typically evaluated in a static framework. The no-anticipation assumption is hidden in the randomization assumption. In a dynamic framework, however, (initial) randomization is not sufficient since evaluation typically conditions on survival. Due to dynamic selection, the composition in the different treatment arms might change differently. Here, assumption 2 is sufficient to complement a static randomization assumption. We establish the link explicitly in the next section.

No anticipation in setup I. Assumption 2 has a subtle additional implication in setup I. Because the spells of treatment and control cohorts begin at different calendar dates, an elapsed duration of t time units is also reached at different dates. Thus, expression (4) not only requires that individuals do not anticipate the (date of the) treatment, but also that the economic conditions of treatment and control cohorts are identical. This requirement can be seen as a stationarity requirement on the data generation process (no cohort effects). As an example, if the local labor market conditions change substantially between two cohorts, then assumption 2 would be violated even if individuals do not anticipate the treatment.

No anticipation in the dynamic literature. In the context of ALMPs, the no-anticipation assumption has been adopted throughout the theoretical and empirical literature on dynamic treatment effects (Sianesi, 2004; Vikström, 2017; Abbring & van den Berg, 2003). Much of this literature is surveyed in Crepon et al. (2018).

No anticipation in the static evaluation literature. Assumption 2 is often implicitly assumed in static evaluation approaches. One example is the phased-in experimental design discussed above. Another example is the DiD approach. Anticipation effects potentially undermine the parallel trends assumption (see Lalive, 2008, for an application in a labor

market context). A third example is the synthetic control approach, where assumption 2 is an implicit component of the conditional independence assumption.

Assumption 3: Randomization. It holds that

$$(i) \quad Z \perp\!\!\!\perp \{T(s), S(t)\}_{t,s \in \mathbb{R}_+ \cup \{+\infty\}} \mid X, V \quad \text{and} \\ (ii) \quad Z \perp\!\!\!\perp V \mid X. \quad (6)$$

Assumption 3 is a randomization assumption. In the context of setup II, assumption 3i postulates that assignment to treatment (Z) is independent of potential outcomes conditional on observed and unobserved covariates. Since X and V are assumed to fully describe an individual, this is an innocuous assumption. The major component of this is assumption 3ii. It states that the instrument Z is independent of unobservables conditionally on observables. It holds the following implication:

$$A3(i), A3(ii) \Rightarrow Z \perp\!\!\!\perp \{T(s), S(t)\} \mid X. \quad (7)$$

Thus, under assumption 3, assignment of Z is driven by X .

Example IIa (continued): Suppose that a case worker is responsible for a large pool of unemployed individuals. Then it can be argued that she acts idiosyncratically given (objective) characteristics of the unemployed and her own assessment of the unemployed.¹³ In such cases, assumption 3 is valid if both the objective characteristics and the case worker's assessment are also available to the econometrician. As a further example, consider randomized phased-in experiments. Assumption 3 is valid per design construction. In the case of phased-in implementation of social programs, this assumption holds whenever the early or late recipients are randomly selected. In setup I, the plausibility of assumption 3 hinges on the stability of the economic environment: it requires identical economic environments at the dates of inflow of young and old cohorts. The relation to assumption 3 in setup II is best explained with a thought experiment, which features individuals randomly assigned to different cohorts. Differences in the structural economic parameters of treated versus non-treated cohorts (cohort effects) violate assumption 3. Such differences can be caused by, for example, mass layoffs and macroeconomic trends.

Assumption 4: Consistency. For all $t, s \in \mathbb{R}_+ \cup \{+\infty\}$:

- i. $Z = t \Rightarrow S(t) = S.$
- ii. $S = s \Rightarrow T(s) = T.$

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

¹³As an example, in Sianesi (2004), such assessments relate to the job seeker's degree of job readiness, as well as to the job seeker's inclinations and urgency to find a job. These assessments are documented and part of the observed covariates used in Sianesi (2004).

potential outcomes and observables. Assumptions 1 and 4 imply together that the actual elapsed duration at which the treatment is received, S , can be either equal to Z or to ∞ .

In addition to assumptions 1 to 4, we implicitly assume that all expressions below exist. This amounts to common support assumptions such as $0 < P(S = t \mid X, V, Z = t)$. These assumptions are fulfilled, for example, when S and Z are discrete, but it is sufficient that S and Z have a positive probability mass on t and t' . Whether discrete Z and S impose a restriction on the distribution of T depends on the concrete application. In the medical treatment example, a specific therapy might be assigned only at predetermined, common-for-everybody, elapsed-time intervals of the disease, whereas the life or disease duration itself is a continuous variable. In the labor market example, the administrative duration of unemployment is always discrete. Nevertheless, it is usually modeled in the literature as a continuous variable, especially when it is measured daily. Labor market treatments such as training and counseling measures or financial penalties might be designed to come into force only at coarser time intervals. Therefore, it might be practical to model them as discrete variables.

Remark 6: Potential Biases in Setup I.¹⁴ Denote the calendar date of treatment in setup I with day 0. Consider cohorts $Z = t$ and $Z = t'$, which enter unemployment at dates $-t$ and $-t'$, respectively, with $t < t'$. From the discussion of assumptions 2 and 3, nonstationary economic environment emerges as a possible source of bias in setup I. We distinguish between two cases. In case 1, the economic environment is stationary up to date 0, and there is a structural change after 0. This can be the case, for example, when other economic policies are implemented at 0 alongside the ALMP of interest. One approach to isolate the effect of the ALMP of interest would be to consider the limit case $a \rightarrow 0$. The validity of this approach would hinge on the assumption that the change in the economic environment at or after 0 has no instantaneous effect on the unemployment duration. In this case, our main identification strategy, described in section IIIB, would yield an unbiased estimator of the treatment effect on the hazard. In case 2, the general case of different economic conditions, the hazard approach is not sufficient. One approach would be to consider the limit case $t' \rightarrow t+$, where the $+$ sign indicates convergence from the right. The intuition of this approach is that individuals who enter unemployment at two points in time that are very close are basically evaluated in the same economic environment. This strategy will lead to an RD-type estimator (see section IIIC). A drawback of both approaches (cases 1 and 2) is the limitation of the set of treatment effects we can evaluate. In both cases, we restrict the window a to be very short.¹⁵ This drawback is closely related to the critique of the phased-in design experiments, in which eval-

uation of long-term effects is in general not possible (Duflo et al., 2007).

B. Identification Results

Assume first that T is observable. Consider expression (2). The main challenge is to find a control group for those who survived until t and were treated at t . In particular, $T(t')$ and $S(t)$ are never jointly observed. They correspond to different treatments (t' and t). Thus, one of the outcomes, $T(t')$, $S(t)$, is always counterfactual. To motivate our identification strategy, consider the following naive candidates for a treatment effect:

$$P(T \in [t, t+a] \mid T \geq t, X, S = t, Z = t) - P(T \in [t, t+a] \mid T \geq t, X, S = \infty, Z = t), \quad (8)$$

$$P(T \in [t, t+a] \mid T \geq t, X, S = t, Z = t) - P(T \in [t, t+a] \mid T \geq t, X, S = t', Z = t'), \quad (9)$$

for $t' > t$. For simplicity, we set the discussion in the context of setup I. Writing equation 8 in the form

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

makes it clear that equation (8) compares averages over two different subgroups of the same cohort: the t -compliers and the t -noncompliers. Since S is a choice variable, it holds in general that

$$V \perp\!\!\!\perp S \mid T \geq t, X, Z = t. \quad (10)$$

Thus, equation (8) captures not only the treatment effect but also the selection bias. The unsurprising implication is that noncompliers are not suitable as a control group.

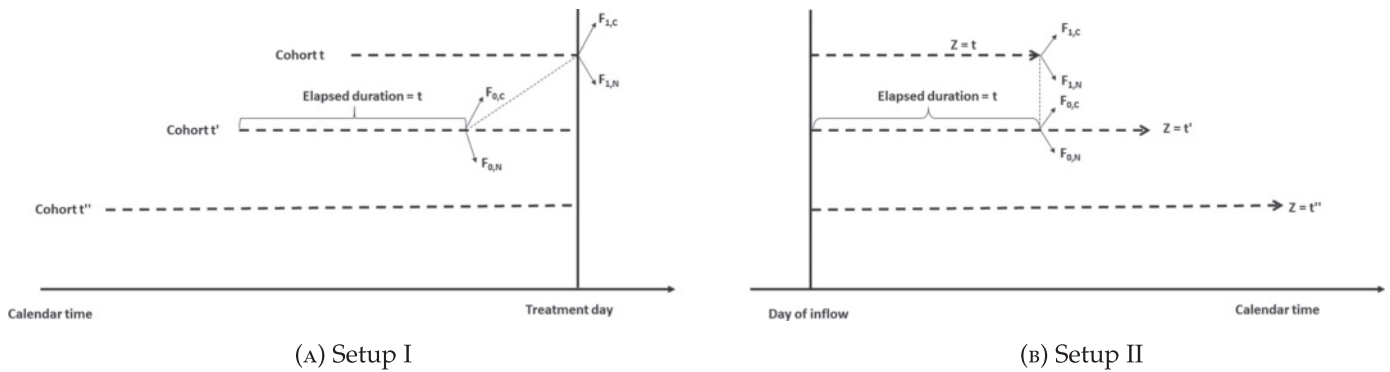
Expression (9) compares the average outcome at elapsed duration t of the t -compliers from the younger cohort $\{Z = t\}$ with the average outcome (at the same duration) of the t' -compliers from the older cohort $\{Z = t'\}$. In general, however, $F_{V \mid T \geq t, S=t, Z=t} \neq F_{V \mid T \geq t, S=t', Z=t'}$ due to dynamic selection. This follows because some unemployed might find a job between elapsed durations t and t' , while others might change their preferences. As a result, learning the compliance status at t' is also not helpful for constructing a control group. The above considerations apply equivalently for setup II.

Instead, our strategy combines the approach of the dynamic treatment effects literature with the static LATE approach. Thus, identification consists of two steps: a dynamic and a static one. Our dynamic step, presented in the next

¹⁴This remark and the resulting RD analysis in section IIIC result from a hint by an anonymous referee, for which we are thankful.

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

FIGURE 2.—IDENTIFICATION IN SETUPS I AND II



proposition, extends the result of van den Berg et al. (2020) to a setting with endogenous noncompliance.¹⁶

Proposition 1. *Let F be a cdf. Under assumptions 1 to 4, it holds for all $\infty \geq t' \geq t \geq 0$:*

$$F_{V|T(t) \geq t, X, S(t)=t} = F_{V|T(t') \geq t, X, S(t)=t} = F_{V|T \geq t, X, S=t, Z=t} \quad \text{and} \quad (11)$$

$$F_{V|T(t) \geq t, X, S(t)=\infty} = F_{V|T(t') \geq t, X, S(t)=\infty} = F_{V|T \geq t, X, S=\infty, Z=t}. \quad (12)$$

To interpret proposition 2, consider example I. Assume that we observe two cohorts, t and t' , of unemployed individuals with dates of inflow $-t = -6$ and $-t' = -9$ (0 is set to be equal to 01.07.2001, the day of the implementation of the reform). The first equality in equation (11) states that under assumptions 1 to 4, individuals in cohort t who have been unemployed for at least six months and are willing to take the treatment have the same distribution of V as individuals in cohort t' who also have been unemployed for at least six months and are willing to take the treatment. The second equality of equation (11) links potential to observed conditions. Equation (12) provides an equivalent result for the group of t -noncompliers. Note that an implication of proposition 2 is that the treatment effects on the treated survivors, $\{T(t) \geq t\}$, and on the untreated survivors, $\{T(t') \geq t\}$, and hence on all survivors, coincide.

The intuition behind proposition 2 is the following. If the unobserved heterogeneity V has the same distribution in the two cohorts at the point in time of inflow (conditional on X) and if these distributions evolve over time in the same way, then V will have the same distribution in the two cohorts at a later pretreatment elapsed duration $t > 0$ (see the dotted lines in figures 2a and 2b). The equality of the distributions of V at $t = 0$ is ensured by the randomization assumption 3. The dynamics is controlled by the “no anticipation” assumption 2. The interpretation and intuition for setup II are equivalent with cohorts replaced by groups of individuals who are assigned to the treatment at different dates.

To motivate the second step of our approach (the static step), consider the following corollary of proposition 2.

Corollary 1. *Let $a \leq t' - t$. Under assumptions 1 to 4, it holds for all $\infty \geq t' \geq t \geq 0$ that*

$$TE(t, t', a) = P(T(t) \in [t, t + a] | T(t) \geq t, X, S(t) = t) - P(T(t') \in [t, t + a] | T(t') \geq t, X, S(t) = t). \quad (13)$$

The right-hand-side of equation (13) does not contain the unobserved V . Thus, if the compliance status ($S(t)$) was observed, the treatment effect could be calculated by comparing the average outcome of compliers from the treated group $\{Z = t\}$, denoted by $F_{1,C}$, with the average outcome of compliers from the not-yet-treated group, $\{Z = t'\}$, denoted by $F_{0,C}$. For the subgroup $\{Z = t\}$, $S(t)$ is observed at elapsed duration t . Therefore, $F_{1,C}$ is identified. For the not-yet-treated $\{Z = t'\}$, however, $S(t)$ is unobserved.

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

$$F_0 = F_{0,C}P_{0,C} + F_{0,N}P_{0,N}, \quad (14)$$

where $F_{0,N}$ is the average outcome for noncompliers in the untreated group, and $P_{0,C}$ and $P_{0,N}$ are the proportions of compliers and noncompliers in that group, respectively (all at elapsed duration t). Figures 2a and 2b illustrate the idea.

Solving equation (14) for $F_{0,C}$ yields $F_{0,C} = (F_0 - F_{0,N}P_{0,N})/P_{0,C}$. Hence, in order to identify $F_{0,C}$, it is sufficient to identify F_0 , $P_{0,C}$, $P_{0,N}$, and $F_{0,N}$. We can directly link F_0 to observable outcomes at elapsed duration t ; it is equal to the average observed outcome at t of the whole not-yet-treated group, $P(T \in [t, t + a] | T \geq t, Z = t')$. Thus, F_0 is identified. In addition, under assumptions 1 to 4, we can identify $P_{0,C}$, $P_{0,N}$, and $F_{0,N}$ from the treated group. In particular, due to randomization and no-anticipation, all pretreatment characteristics of the two groups have equal distributions at elapsed duration t . It holds therefore that $P_{0,C} = P_{1,C}$, $P_{0,N} = P_{1,N}$, and $F_{0,N} = F_{1,N}$. The intuition for the last equality is that in both groups, noncompliers are not treated. The

¹⁶All proofs are in the online appendix.

exclusion restriction, equation (1), ensures that the assignment to treatment alone does not change their outcomes.

With these considerations, we can state our main identification result:

Proposition 2. *Let $a \leq t' - t$. Under assumptions 1 to 4, the treatment effect on the treated $TE(t, t', a)$ is nonparametrically identified for all $\infty \geq t' \geq t \geq 0$ and it holds that*

$$TE(t, t', a) = \frac{\left(\begin{array}{l} P(T \in [t, t + a] | T \geq t, X, Z = t) \\ - P(T \in [t, t + a] | T \geq t, X, Z = t') \end{array} \right)}{P(S = t | T \geq t, X, Z = t)}. \tag{15}$$

Expression (15) has an intuitive interpretation. It adjusts the difference between the average observed outcomes in the two groups by the probability of being a complier. The adjustment takes account of the fact that any difference between the outcomes of the two groups can be caused only by the compliers. Our result is related to the static one-sided non-compliance result of Bloom (1984) and the LATE identification result in Imbens and Angrist (1994). Identification is local in the sense that the treatment effect is identified only for the subgroup of t -compliers. As this group is allowed to change with t , our notion of location can be seen as a dynamic extension of the LATE notion of location.

We now consider the case of right-censoring. In a labor market context, right-censoring typically arises when, at the end of the period of observation, some individuals are still unemployed, so their unemployment spells have an unknown length. Censoring occurs also when unemployment is interrupted by a transition out of the labor force due to maternity, sickness, or military service, or simply when individuals do not show up to report about their status (attrition). With a small abuse of notation, let $C \geq 0$ be the censoring r.v. Define $\tilde{T} = \min\{T, C\}$ and $\delta = \mathbb{1}\{\tilde{T} = T\}$. We observe (\tilde{T}, δ) and, not directly, (T, C) . It is not possible to recover nonparametrically the joint distribution of T and C from the distribution of (\tilde{T}, δ) without additional assumptions (see Tsiatis, 1975). We adopt the following additional standard assumption:

Assumption 5: Random censoring. $C \perp (T, S) | X, Z$.

It holds the following proposition.

Proposition 3. *Under assumptions 1 to 5, $TE(t, a)$ is identified.*

Remark 7. Identification of a treatment effect on the hazard (HTE) requires only a mild additional regularity assumption and is left for the online appendix.

Remark 8. Under assumptions 1 to 4, we have $P(T \in [t, t + a] | T \geq t, X, Z = t') = P(T \in [t, t + a] | T \geq t, X, Z = t'')$ for all $t', t'' \geq t + a$ (in the limit case $a \rightarrow 0$ simply for $t', t'' > t$). As a consequence, the treatment effect does not depend on the choice of the nontreated cohort t' as long as $t' \geq t + a$ (or $t' > t$). Therefore, we omit the dependence on t' and write $TE(t, a)$ and $HTE(t)$.

C. Identification in a Regression Discontinuity Setup

In this section, we outline an identification approach that mitigates the problems related to setup I discussed in remark 6.¹⁷ The intuition is that if t' is sufficiently close to t , $t < t'$, then treatment and control cohorts are practically evaluated in the same economic environment. This consideration leads to the following modifications of assumptions 2 and 3:

Assumption 2'. $P(T(t) > t | X) = \lim_{t' \rightarrow t+} P(T(t') > t | X)$,

Assumption 3'. Denote by $F_{T(s), S(z) | Z, X, V}$ and $F_{V | Z, X}$ the corresponding conditional distributions. Then there exists a positive number η , such that for all t' in the η -neighborhood of t and for all positive (s, z)

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

Assumptions 2' and 3' are local versions of assumptions 2 and 3, respectively. Note that assumption 2' is almost trivially fulfilled in both empirical setups. In particular, even if an individual has perfect knowledge of the assigned treatment (t or t'), the effect of this knowledge under t and t' will be the same when $t' \rightarrow t+$.

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

Proposition 4. *Assume that $\lim_{t' \rightarrow t+} \theta(t | X, Z = t')$ exists. Then, under assumptions 1, 2', 3', 4, and 5 and HTE1, the treatment effect on the hazard at t is identified and equal to*

$$\Psi_{RD}(t) = \lim_{t' \rightarrow t+} \frac{\theta(t | X, Z = t) - \theta(t | X, Z = t')}{P\{S = t | T \geq t, X, Z = t\}}. \tag{16}$$

Equation (16) is related to the static regression discontinuity (RD) identification result in Hahn et al. (2001). Z can be seen as a forcing variable with a discontinuity at t . Assumption 3' is related to the conditional independence assumption made in Theorem 2 in Hahn et al. (2001). Assumption 2', however, is new and needed to account for dynamic selection. Thus, an estimator of the treatment effect based on assumptions 2' and 3' and result (16) can be interpreted as a dynamic RD estimator.

A practical problem of this approach is that the finite sample performance of the estimator might lack precision due to a lack of observations near the boundary. In addition, if the treatment effect does not substantiate instantaneously, it will not be detected by the estimator.

¹⁷We are thankful to an anonymous referee for suggesting this strategy to us.

IV. Estimation

In this section, we develop an estimator for $TE(t, a)$. An estimator for the treatment effect on the hazard follows van den Berg et al. (2020) and is presented in section A.2 in the appendix. For simplicity of exposition, we ignore the dependence on covariates. All results that follow generalize in a straightforward way to the case with covariates. One simply uses the conditional Kaplan-Meier estimator of Gonzalez-Manteiga and Cadarso-Suarez (2007) instead of its unconditional counterpart.

Let $t < t'$. Define $\bar{F}_1(t) = P\{T > t \mid Z = t\}$, $\bar{F}_2(t) = P\{T > t \mid Z = t'\}$, and $p = P\{S = t \mid T \geq t, Z = t\}$. The former two are observed survival probabilities and the latter is the compliance probability. Under assumptions 1 to 4, $TE(t, a)$ can be written as

$$TE(t, a) = \frac{1}{p} \left(\frac{\bar{F}_2(t+a)}{\bar{F}_2(t)} - \frac{\bar{F}_1(t+a)}{\bar{F}_1(t)} \right). \tag{17}$$

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

$$\hat{F}_j(t) = \bar{F}_j(t) + o_p(1), \tag{18}$$

$$\sqrt{n} \left(\hat{F}_j(t) - \bar{F}_j(t) \right) \xrightarrow{d} N(0, \sigma_j^2(t)) \text{ as } n \rightarrow \infty, \tag{19}$$

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

$$\widehat{TE}(t, a) = \frac{1}{\hat{p}} \left(\frac{\hat{F}_2(t+a)}{\hat{F}_2(t)} - \frac{\hat{F}_1(t+a)}{\hat{F}_1(t)} \right). \tag{20}$$

Equation (20) can be interpreted as a dynamic version of a Wald estimator. Its consistency is stated in the following proposition.

Proposition 5. *Suppose equation (18) holds. Then under assumptions 1 to 5, it holds that*

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

The following proposition states the asymptotic distribution of the estimator.

Proposition 6. *Let assumptions 1 to 5 and condition (19) hold. Then it holds*

$$\begin{aligned} \sqrt{n}(\widehat{TE}(t, a) - TE(t, a)) &\xrightarrow{d} N\left(0, \frac{1}{p^2} \sum_{i=1}^2 \left(\frac{1}{\bar{F}_i^2(t)} \sigma_i(t+a) \right. \right. \\ &\left. \left. + \frac{\bar{F}_i^2(t+a)}{\bar{F}_i^4(t)} \sigma_i(t) + \frac{\bar{F}_i(t+a)}{\bar{F}_i^3(t)} \sigma_i(t, t+a) \right) \right), \end{aligned} \tag{21}$$

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

Confidence bands can be constructed by replacing the unknown terms in the variance with consistent estimates, for example using the Greenwood's formula (Andersen et al., 1997). It follows from equation (21) that the precision of the estimator is inversely related to p . The bigger the compliance probability p (i.e., the stronger the instrument Z), the smaller the variance.

V. Model Diagnostics

A. Testing the Assumptions

We suggest two testing approaches to address the main assumptions, 2 and 3.

The first approach parallels model diagnostics in a DiD context and focuses on pretreatment outcomes. Empirical tests for equality of pretreatment survival probabilities are applied by De Giorgi (2005) and van den Berg et al. (2020). In this section, we give a theoretical justification for these tests. Let $s \leq t < t'$. Under assumptions 1 and 4, assumptions 2 and 3 jointly imply

$$P(T > s \mid X, Z = t) = P(T > s \mid X, Z = t'). \tag{22}$$

In words, observed survival probabilities are equal in the treated and nontreated groups at any common elapsed pretreatment duration. Adopting assumptions 1 and 4 as fundamental assumptions, equality (22) can be used to test for no anticipation and randomization. This is captured in the following proposition (dependency on X is ignored).

Proposition 7. *Let assumptions 1, 4, and 5 and condition (19) hold. Then, under the null hypothesis H_0 : assumption 2 \cup assumption 3, it holds that*

$$\sqrt{N}(\hat{F}_1(t) - \hat{F}_2(t)) \overset{d}{\sim} N(0, 2\sigma^2(t)). \tag{23}$$

The second approach is to adopt the weaker assumptions, 2' and 3' (together with assumptions 1, 4, and 5) as fundamental and use equation (16) in a DiD setup. Let $s > t$. Then under assumption 2 and 3, the expression

$$\begin{aligned} &(\theta(t \mid X, Z = t) - \theta(t \mid X, Z = s)) - (\theta(t \mid X, Z = t) \\ &- \lim_{t' \rightarrow t+} \theta(t \mid X, Z = t')) = \lim_{t' \rightarrow t+} \theta(t \mid X, Z = t') \\ &- \theta(t \mid X, Z = s) \end{aligned} \tag{24}$$

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

B. Framework for the Analysis of Endogeneity

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

$$(i) F_{0,C} = F_{0,N} \quad \text{and} \quad (ii) F_{1,C} = F_{1,N}, \quad (25)$$

where we used the simplified notation from section IIIB. It follows from assumption 1 that $F_{1,N}$ is not identified: t -noncompliers are never observed under the treatment t . As a result, hypothesis (25ii) is not testable. We therefore focus on hypothesis (25i). It follows from relation (14) that hypothesis (25i) is equivalent to $F_0 = F_{0,N}$. In particular, if compliers and noncompliers have equal (average) potential outcome distributions, then the outcome distribution of the whole population under no treatment is equal to the outcome distribution of the noncompliers under no treatment. Furthermore, we showed in section IIIB that F_0 and $F_{0,N}$ are identified under assumptions 1 to 4, with

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

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

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

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

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

The bias that arises from an endogenous non-take-up can be measured as the difference between the true treatment effect (15), $TE = F_{1,C} - F_{0,C}$, and the naive candidate (8) (for short, NTE), $NTE = F_{1,C} - F_{1,N}$ (dependence on t and a is suppressed). Define $\mathbb{B} = TE - NTE$. Substituting $F_{0,C} = (F_0 - F_{0,N}P_N)/P_C$ and $F_{0,N} = F_{1,N}$ yields $\mathbb{B} = (F_{0,N} - F_0)/P_C = \mathbb{D}/P_C$. An empirical analysis of the bias

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

$$\tilde{H}_0: P(T \geq t \mid X, Z = t') - P(T \geq t \mid X, S = \infty, Z = t) = 0. \quad (26)$$

from endogenous selection can be performed with an estimator of \mathbb{B} .

VI. Empirical Application: The French PARE Labor Market Reform in 2001

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

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

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

$$TE(6, a) = P(T(6) \in [6, 6+a] \mid T(6) \geq 6) \\ - P(T(t') \in [6, 6+a] \mid T(t') \geq 6), \quad (27)$$

where $t' > 6$ and $a < t' - t$.²²

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

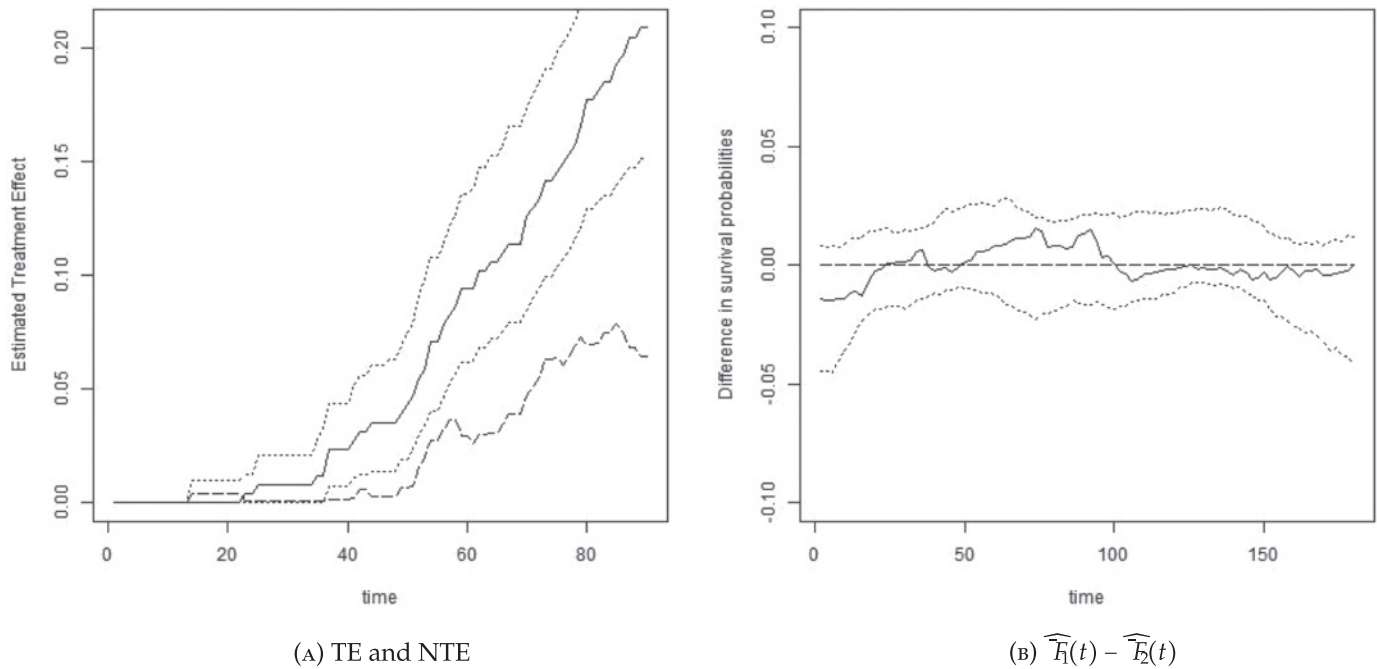
¹⁹Some studies that evaluate the French unemployment insurance system are Fougère, Kamionka, and Prieto (2010); Debauche and Jugnot (2007); Crépon, Ferracci, and Fougère (2012); Crépon, Dejemeppe, and Gurgand (2005); and Le Barbanchon (2012).

²⁰A comprehensive description of the reform can be found in Freyssinet (2002).

²¹We describe the data set and our empirical strategy (and, in particular, the motivation behind our choice of the treated cohort) in section B.1 in the appendix.

²²We use corollary 1 here to simplify the expression for the treatment effect.

FIGURE 3.—EMPIRICAL RESULTS



(a) Estimated Treatment Effect (Thick Line), 95% confidence bounds (dotted line), Naive treatment effect (dashed line). (b) Difference of pretreatment survival functions (thick line), 95% confidence bounds (dotted line), the zero axis (dashed line).

of white- and blue-collar workers and for the subgroups of unemployed with higher and lower education status follow a pattern very similar to the pattern of the unconditional estimates.

Next, we perform thorough model diagnostics. First, chi-square tests for equality of distributions reveal that pretreatment observed characteristics are balanced between the treatment and control groups (see appendix B.3.1). Second, we also provide evidence that the economic conditions at the inflow of the two cohorts were very similar (also in appendix B.3.1). These findings support the plausibility of assumption 3 and, in particular, that the estimates are not undermined by cohort effects. Third, we estimate the TE with an alternative choice of the control cohort ($t'' = 8$ months). The estimation results are very similar to the main results (see figure 3 in the appendix). This is further evidence that there were no cohort effects. In section B.3.2, we provide arguments that the reform was not anticipated by the unemployed due to the relatively short notice and lack of (clear) political debate. Fourth, we test for equality of pretreatment survival functions along the lines of section VA. The estimated difference of the survival functions of treatment and control groups is depicted in figure 3b with a thick line. The 0 axis is fully contained within the 95% confidence bounds. Therefore, we cannot reject the joint hypothesis (assumptions 2 and 3).

Finally, we analyze the non-take-up of the reform (non-compliance) along the lines of section VB. To do so, we compare the IV estimate $\widehat{TE}(6, a)$ with an estimate of the NTE, equation (8) (see the dashed line in figure 3a). At all points at which the TE is positive, the estimated NTE is smaller than

the corresponding estimated TE. The difference is significant at the 95% level.

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

REFERENCES

- Abbring, Jaap H., and Gerard J. van den Berg, "The Non-Parametric Identification of Treatment Effects in Duration Models," *Econometrica* 71 (2003), 1491–1517.
- , "Social Experiments and Instrumental Variables with Duration Outcomes," Tinbergen Institute discussion paper 05-047/3 (2005).
- Andersen, Per K., Ørnulf Borgan, Richard D. Gill, and Niels Keiding, *Statistical Models Based on Counting Processes* (New York: Springer, 1997).
- Biewen, Martin, Bernd Fitzenberger, Aderonke Osikominu, and Marie Paul, "The Effectiveness of Public-Sponsored Training Revisited: The Importance of Data and Methodological Choices," *Journal of Labor Economics* 32 (2014), 837–897. doi:10.1086/677233.
- Bijwaard, Govert E., "Instrumental Variable Estimation for Duration Data," (pp. 111–148), in Henriette Engelhardt, Hans-Peter Kohler, and Alexia Fürnkranz-Prskawetz, eds., *Causal Analysis in Population Studies: Concepts, Methods, Applications* (Berlin: Springer, 2009).
- Bijwaard, Govert E., and Geert Ridder, "Correcting for Selective Compliance in a Reemployment Bonus Experiment," *Journal of Econometrics* 125 (2005), 77–111.

- Blasco, Sylvie, "Do People Forgo Extra Money to Avoid Job Search Assistance?" CREST discussion paper (2009).
- Bloom, Howard S., "Estimating the Effect of Job-Training Programs, Using Longitudinal Data: Ashenfelter's Findings Reconsidered," *Journal of Human Resources* 19 (1984), 544–556.
- Blundell, Richard, Monica Costa Dias, Costas Meghir, and John Reenen, "Evaluating the Employment Impact of a Mandatory Job Search Program," *Journal of the European Economic Association* 2 (2004), 569–606. doi:10.1162/154247604142336.
- Carney, Terry, and Gabi Ramia, "Welfare Support and Sanctions for Noncompliance in a Recessionary World Labour Market: Post-Neoliberalism or Not?" Sydney Law School technical report 11 (2011).
- Chesher, Andrew, "Semiparametric Identification in Duration Models," CeMMAP working paper CWP20/02 (2002).
- Crépon, Bruno, Muriel Dejemepe, and Marc Gurgand, "Counseling the Unemployed: Does It Lower Unemployment Duration and Recurrence?" IZA discussion paper 1796 (2005).
- Crépon, Bruno, Marc Ferracci, and Denis Fougère, "Training the Unemployed in France: How Does It Affect Unemployment Duration and Recurrence?" *Annales d'Economie et de Statistique* 107–108 (2012), 175–199.
- Crépon, Bruno, Marc Ferracci, Gregory Jolivet, and Gerard J. van den Berg, "Active Labor Market Policy Effects in a Dynamic Setting," *Journal of the European Economic Association* 7 (2009), 595–605.
- "Information Shocks and the Empirical Evaluation of Training Programs during Unemployment Spells," *Journal of Applied Econometrics* 33 (2018), 594–616.
- De Giorgi, Giacomo, "The New Deal for Young People Five Years On," *Fiscal Studies* 26 (2005), 371–383. doi:10.1111/j.1475-5890.2005.00016.x.
- Debauche, Etienne, and Stephane Jugnot, "Les effets du projet d'action personnalisé sur les sorties des listes de l'ANPE," DARES working paper (2007).
- Duflo, Esther, Rachel Glennerster, and Michael Kremer, "Using Randomization in Development Economics Research: A Toolkit" (4:3895–3962), in T. Paul Schultz and John A. Strauss, eds., *Handbook of Development Economics* (Amsterdam: Elsevier, 2007). doi:https://doi.org/10.1016/S1573-4471(07)04061-2.
- Eberwein, Curtis, John C. Ham, and Robert J. LaLonde, "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 (1997), 655–682.
- Field, Erica, "Entitled to Work: Urban Property Rights and Labor Supply in Peru," *Quarterly Journal of Economics* 122 (2007), 1561–1602.
- Fougère, Denis, Thierry Kamionka, and Ana Prieto, "L'efficacité des mesures d'accompagnement sur le retour à l'emploi," *Revue Économique* 61 (2010), 599–612.
- Freyssinet, Jacques, "La réforme de l'indemnisation du chômage en France," IRES document de travail 02.01 (2002).
- Gonzalez-Manteiga, Wenceslao, and Carmen Cadarso-Suarez, "Asymptotic Properties of a Generalized Kaplan-Meier Estimator with Some Applications," *Journal of Nonparametric Statistics* 4 (2007), 65–78.
- Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica* 69 (2001), 201–209.
- Ham, John C., and Robert J. LaLonde, "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica* 64 (1996), 175–205.
- Hausman, Jerry A., and Tiemen M. Woutersen, "The Proportional Hazard Model," in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics* (Basingstoke: Palgrave Macmillan, 2008).
- "Estimating a Semi-Parametric Duration Model without Specifying Heterogeneity," *Journal of Econometrics* 178 (2014), 114–131.
- Heckman, James J., "Econometric Causality," *International Statistical Review* 76 (2008), 1–27.
- Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith, "The Economics and Econometrics of Active Labor Market Programs," (3:1865–2097), in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics* (Amsterdam: Elsevier, 1999).
- Heckman, James J., and Salvador Navarro, "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics* 136 (2007), 341–396.
- Imbens, Guido W., and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62 (1994), 467–475.
- Imbens, Guido W., and Donald B. Rubin, "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies* 64 (1997), 555–574.
- Kalbfleisch, John D., and Ross L. Prentice, *The Statistical Analysis of Failure Time Data* (New York: Wiley, 2002).
- Lalive, Rafael, "How Do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach," *Journal of Econometrics*, 142 (2008), 785–806.
- Lalive, Rafael, Jan C. van Ours, and Josef Zweimüller, "The Effect of Benefit Sanctions on the Duration of Unemployment," *Journal of the European Economic Association* 3 (2005), 1386–1417. doi:10.1162/154247605775012879.
- Lancaster, Tony, "Econometric Methods for the Duration of Unemployment," *Econometrica* 47 (1979), 939–956.
- *The Econometric Analysis of Transition Data* (Cambridge: Cambridge University Press, 1990).
- Le Barbanchon, Thomas, "The Effect of the Potential Duration of Unemployment Benefits on Unemployment Exits to Match Quality in France," CREST working paper (2012).
- Lechner, Michael, Ruth Miquel, and Conny Wunsch, "Long-Run Effects of Public Sector Sponsored Training in West Germany," *Journal of the European Economic Association* 9 (2011), 742–784.
- Meyer, Bruce D., "What Have We Learned from the Illinois Reemployment Bonus Experiment?" *Journal of Labour Economics* 14 (1996), 26–51.
- Miguel, Edward, and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica* 72 (2004), 159–217. doi:10.1111/j.1468-0262.2004.00481.x.
- Robins, James M., and Anastasios A. Tsiatis, "Correcting for Non-Compliance in Randomized Trials Using Rank Preserving Structural Failure Time Models," *Communications in Statistics—Theory and Methods* 20 (1991), 2609–2631.
- Sianesi, Barbara, "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," this REVIEW 86 (2004), 133–155.
- Tchetgen, Eric J. T., Stephan Walter, Torben Martinussen, and Maria M. Glymour, "Instrumental Variable Estimation in a Survival Context," Harvard Biostatistics working paper series 179 (2014).
- Todd, Petra E., "Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated" (4:3847–3894), in T. Paul Schultz and John A. Strauss, eds., *Handbook of Development Economics* (Amsterdam: Elsevier, 2007). doi:https://doi.org/10.1016/S1573-4471(07)04060-0.
- Tsiatis, Anastasios, "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proc. Nat. Acad. Sci.* 72 (1975), 20–22.
- van den Berg, Gerard J., "Duration Models: Specification, Identification, and Multiple Durations" (pp. 3381–3460), in J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics* (Amsterdam: Elsevier, 2001).
- van den Berg, Gerard J., Antoine Bozio, and Mónica Costa Dias, "Policy Discontinuity and Duration Outcomes," *Quantitative Economics* (2020, Forthcoming).
- Vikström, Johan, "Dynamic Treatment Assignment and Evaluation of Active Labor Market Policies," *Labour Economics* 49 (2017), 42–54. doi:https://doi.org/10.1016/j.labeco.2017.09.003.