INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2010-41

Analyzing the Anticipation of Treatments with Data on Notification Dates

B. CRÉPON[1] – M. FERRACCI[2]
G. JOLIVET[3] – G. van den BERG[4]

[1] CREST-INSEE. Email : crepon@ensae.fr

[2] Université Paris-Est, ERUDITE and CREST-INSEE. Email : marc.ferracci@ensae.fr

[3] University of Bristol Corresponding author : gregory.jolivet@bristol.ac.uk

[4] University of Mannheim, VU University Amsterdam, IFAU and IZA. Email : gjvdberg@xs4all.nl

Analyzing the Anticipation of Treatments with Data on Notification Dates*

Bruno Crépon† **Marc Ferracci**‡ **Grégory Jolivet**§
Gerard J. van den Berg¶

July 2010

Abstract

When treatments may occur at different points in time, most evaluation methods assume - implicitly or explicitly - that all the information used by subjects about the occurrence of a future treatment is available to the researcher. This is often called the "no anticipation" assumption. In reality, subjects may receive private signals about the date when a treatment may start. We provide a methodological and empirical analysis of this issue in a setting where the outcome of interest as well as the moment of information arrival (notification) and the start of the treatment can all be characterized by duration variables. Building on the "Timing of Events" approach, we show that the causal effects of notification and of the treatment on the outcome are identified. We estimate the model on an administrative data set of unemployed workers in France which provides the date when job seekers receive information from caseworkers about their future treatment status. We find that notification has a significant and positive effect on unemployment duration. This result violates the standard "no anticipation" assumption and rules out a "threat effect" of training programs in France.

JEL codes: C31, C41, J64, J68

*We thank Sylvie Blasco, Per Johansson, Frank Windmeijer, as well as participants in seminars at ASB Aarhus, the University of Paris I and at the 7^{th} meeting of the EU-RTN "Microdata" project in London, the DARES conference on evaluation in Paris, and EALE/SOLE 2010, for helpful comments. We also thank the Direction of Statistics of Pôle Emploi for providing us with the FHS data.

†CREST-INSEE. E-mail: crepon@ensae.fr

‡Université Paris-Est, ERUDITE and CREST-INSEE. E-mail: marc.ferracci@ensae.fr

§Corresponding author: University of Bristol. E-mail: gregory.jolivet@bristol.ac.uk

¶Alexander von Humboldt Professor, University of Mannheim; VU University Amsterdam, IFAU and IZA. E-mail: gjvdberg@xs4all.nl

1 Introduction

Treatment evaluation problems are often of a dynamic nature. For instance, one may be interested in knowing how the duration an individual spends in a state of interest (say, unemployment) is affected by the moment at which he receives a given treatment (say, training). Over the last fifteen years, new techniques have been developed for the analysis of dynamic treatments. The statistical literature has taken the standard static evaluation framework with potential outcomes, conditional independence, and selection on observables (the Rubin model, 1974) to dynamic discrete-time settings (see Robins, 1997, Lechner, Miquel and Wunsch, 2004, Fredriksson and Johansson, 2008, and Crépon, Ferracci, Jolivet and Van den Berg, 2009). Eberwein, Ham and LaLonde (1997) develop a bivariate discrete duration model where both the treatment and the outcome are duration variables and allow for selection on unobservables. An exclusion restriction is used to identify the causal effect of interest. Abbring and Van den Berg (2003a, henceforth AVdB) prove identification of a continuous-time bivariate duration model with selection on unobservables but without instrumental variables, by exploiting variation in the timing of the treatment versus the outcome. This approach is referred to as the "Timing of Events" approach. A common feature of all these approaches is that they hinge on a crucial assumption which we will refer to as the "no anticipation" (NA) assumption.

In words, the NA assumption states that "the future cannot cause the past", i.e. that an individual's potential outcomes do not depend on future treatments. In our empirical application, the NA assumption implies that the probability that an individual leaves unemployment today is the same whether he will enter a training program tomorrow or next year. As emphasized by AVdB and Abbring and Heckman (2008), it is useful to interpret this assumption in terms of information accumulation over time. If the individual's information set relevant to the future treatment status is fixed over time then inference can proceed in the usual way. In the Timing of Events setting, if this information set is identical for individuals with identical characteristics, then information accumulation may be captured by the model specification to be estimated (see AVdB).

However, if individuals receive at random dates some information shocks that are unobserved by the econometrician, and if they act on the new information, then the NA assumption may be violated. Such a violation of the NA assumption is often plausible. For instance, in the case of active labor market policies, the caseworker may inform the unemployed worker that he has been assigned to a particular treatment (like a training course) that is likely to start within a few weeks. Individuals may act on this information and either wait for the treatment to begin (unemployed workers may stop searching for jobs if they are about to enter a training program) or try to avoid the treatment (unemployed workers may take any job offer in order not to be locked in a training program for several weeks).

As mentioned by AVdB, if the arrival of information is observed by the econometrician, then one way to circumvent the NA assumption is to redefine the problem as an evaluation of the causal effect of the arrival of information.[1] Ideally, one would like to be able to evaluate both the arrival of information and the actual treatment. This is what the present paper sets out to achieve.

Specifically, in this paper, we consider the case where the arrival of information is observed, and we address the full evaluation problem from a methodological and an applied perspective. First, we extend the Timing of Events approach to allow for the arrival of notification shocks that may influence the outcome before the treatment starts. Then, we turn to an empirical application where we use information on notification dates from an administrative data set on unemployed workers in France to test the NA assumption and run an evaluation of training programs when the NA assumption may not hold.

We extend the bivariate duration model adopted in the Timing of Events approach to account for the arrival of a shock that provides individuals with (more) information on their future treatment status. We model the process ruling the arrival of these notification shocks in a similar fashion as those ruling treatment and exit dates. Hence, as will be clear from the presentation of the model in section 2, our approach basically consists in adding one layer to the standard AVdB model. As motivated by our empirical application, we allow for individuals to be treated without notification and for treatment dates to be stochastic conditional on notification. The three processes at play may be interrelated through individual observed and unobserved heterogeneity. Assuming a mixed-proportional structure for the hazard rates, as do AVdB, we can identify the distribution of unobserved heterogeneity from the competing-risks part of the model. Any further correlation between the three durations of interest can then be interpreted as causal. We model four different effects: the effect of notification on treatment, the effect of notification on exit, the effect of treatment with no notification on exit and the effect of treatment preceded by notification on exit. We show that these treatment effects are identified and provide an additional result stating that identification of the effects of notification on treatment and exit can be achieved without

[1] An important alternative approach is developed by Heckman and Navarro (2007). They build on the dynamic discrete-choice literature to propose a discrete-time reduced-form model identified without the NA assumption. This approach requires variation in period-specific instrumental variables, and some exogeneity in the arrival of information shocks. Yet another alternative is to treat information shocks as individual time-varying unobserved heterogeneity. AVdB show that their model can be extended to account for this by using multiple-spell data. However, this may be difficult in practice, as the time needed to gather data on multiple spells may put into question other features of the AVdB model, in particular the stationarity assumption implicit in the mixed-proportional specification of the hazard rates.

imposing any specification on the effects of treatment (with and without notification) on exit. Our test of the NA assumption will be based on the direct effect of notification on exit. If this effect is found to be significant, then notification acts as an information shock that affects individuals' future before they are treated and the NA assumption is violated.

Contrary to the usual case, we allow potential outcomes to differ in the time interval prior to treatment. However, it is intuitively plausible that the addition of an "information arrival" layer below the usual evaluation framework implies that a new NA assumption is required concerning the moment at which information about the future treatment arrives. Specifically, we have to assume that an individual's potential outcomes do not depend on future notification, and that notified individuals' potential outcomes do not depend on future treatments. We thus assume that notification is the only source of variation in individuals' information set. Lastly, we follow the vast majority of the treatment literature and assume that an individual's outcome is not affected by another individual's notification or treatment status.[2]

We take our model to a French administrative register data set that contains all unemployment and training spells experienced by workers in the greater Paris area (Ile-de-France) between 2002 and 2005. In addition to reporting the dates when each worker enters/leaves unemployment or, where relevant, participates in a training program, this data set contains the date when a worker is informed by a caseworker that he is put in contact with a training provider, which is the first step towards state-provided training. We use this information as the notification date. It is important to mention that in France unemployed workers can start a training program without having been notified by caseworkers (in particular when they take the initiative of training). Also, workers who receive notification from a caseworker do not necessarily face sanctions on their unemployment benefits if they shun away from training (they do in theory but sanctions are almost never taken in practice). In this empirical application, we are interested in the effect of notification on exit from unemployment, the effect of notification on training (to a lesser extent) and lastly the effect of training on exit from unemployment. This latter effect may depend on whether the training program was preceded by notification or not.

A range of existing empirical papers study closely related topics. Some studies exploit unusually rich data or institutional settings in order to examine the extent to which individuals adjust their behavior in response to knowledge about the moment of future treatments. Lalive, Van Ours and Zweimüller (2005) use Swiss data that contain the moment at which

[2]This is the stable unit treatment value assumption introduced by Neyman (1923) that rules out equilibrium effects of the treatment. See Ferracci, Jolivet and Van den Berg (2010) for an empirical analysis of this assumption.

the public employment service warns unemployed individuals that they will receive a benefits sanction before it is actually implemented, and they show that this warning increases the propensity to leave unemployment.[3] For this purpose they estimate an extension of the model of AVdB. Our study differs from Lalive et al. (2005) in several aspects, apart from the fact that the two studies analyze completely different treatments and notifications. In our application, notification does not mechanically lead to a treatment or a sanction. Also, whereas in Lalive et al. (2005) warnings always precede sanctions, our model allows for individuals to be treated without having received notification. Lastly, we take a closer look at the identification of our model and show that the effect of notification can be identified without putting any structure on the effect of training. Cockx and Dejemeppe (2007) use a regression discontinuity approach to show that the notification of job search monitoring exerts a "threat effect" in Belgium as well. De Giorgi (2005) and Van den Berg, Bozio and Costa Dias (2010) study the effects of the announcement that a "job search assistance" treatment will take place at a fixed future date on the probability of moving to employment, by comparing a situation where individuals become aware of the moment of the future treatment to situations where the policy regime has not yet been introduced. In all these cases, anticipation of the future treatment is found to affect the outcome.[4]

Our paper also contributes to the empirical literature evaluating the effect of training programs for unemployed workers. This issue has been at the center of the evaluation literature in economics (see Heckman, Lalonde and Smith, 1999, for a survey). Recent evaluations incorporate dynamics (Sianesi, 2004, Crépon et al., 2009, Richardson and Van den Berg, 2006). In France, a recent study by Crépon, Ferracci and Fougère (2007) using the Timing of Events approach finds that training programs have little effect, if any, on exit from unemployment. However, modeling the duration of the subsequent employment spell, they find that the recurrence of unemployment is reduced by training. Our study will enrich the analysis of training programs by providing results on the effect of notification and by showing whether the effect of training depends on notification.

[3]See also Arni, Lalive and Van Ours (2009) who study the effect of warnings and sanctions on employment outcomes.

[4]There is a strong similarity between, on the one hand, the anticipation of treatments due to the arrival of information, and, on the other hand, the behavioral impact of knowing that there is a probability of a future treatment. A range of empirical studies addresses the latter by way of a comparison to a regime where treatments are known to be absent or to have a different probability of occurrence. Black et al. (2003) show that unemployed individuals less often enter an unemployment insurance spell if they learn that this includes compulsory job search assistance. The "threat effect" of compulsory programs has also been found in Danish data by Geerdsen (2006), using policy changes in the duration of benefits for identification, and by Rosholm and Svarer (2008). Van den Berg, Bergemann and Caliendo (2009, 2010) show that newly unemployed workers in Germany have widely different expectations on the probability of future participation in active labor market policies, including training programs, and that this is reflected in their job search behavior as captured by search effort and the reservation wage.

The outline of this paper is as follows: section 2 gives a formal presentation of the NA assumption, describes our model, the effects of notification and treatment, and discusses identification. Section 3 presents the institutional French setting, the notification and training processes, our administrative data, and the econometric specification we use for estimation. All the results are in section 4. Section 5 concludes.

2 Theoretical framework

2.1 The role of the "no-anticipation" assumption in treatment evaluation

We want to evaluate the effect of a treatment on the duration an individual spends in a state of interest. The treatment can be assigned at different points in time. We let Z denote the duration before treatment and Y the duration in the state of interest (also called the outcome). In our empirical application, Y will be the duration in unemployment and Z the duration before the unemployed job seeker enters a training program. The evaluation of the treatment effect builds on a causal model $\left(\{Y(z)\}_{z>0}, Z\right)$, where $Y(z)$ is the potential duration when treatment is received at date $t = z$.[5] Two obstacles stand in the way of identifying the $z \mapsto Y(z)$ functions. First, we observe $Y(Z)$ but do not know anything about $Y(z), z \neq Z$. This means that if an individual has received treatment after, say, 6 months, we cannot predict what would have happened had he received treatment after 1 month , 3 months, 12 months, etc... This is the standard issue of selection into treatment (see Rubin, 1974), which can be overcome by using a relevant set of individual characteristics to write an unconfoundability assumption.

The second issue is related to the dynamic nature of the problem: Z is censored by Y i.e. if the individual leaves the state of interest before receiving treatment, we do not observe Z, we just know that $Z > Y$. This poses a fundamental problem as, if $Z > Y$, we no longer know which process among $\{Y(z)\}_{z>Y}$ corresponds to the observed duration Y. This hampers identification of a causal model unless we put some structure on the missing part of the data. Abbring and Van den Berg (2003a) show that it is necessary to assume that the outcome follows a single process before treatment occurs. This "no-anticipation" assumption can be formally written as follows:[6]

$$\Pr\left(Y(z) = t\right) = \Pr\left(Y(z') = t\right), \quad \forall t \leq \min(z, z'). \tag{1}$$

[5] To simplify the presentation, we assume that the individual enters the state of interest at date $t = 0$.

[6] The NA assumption should be made conditional on the set of confounders used to solve the first identification issue. For expositional convenience, we do not write this conditioning in this subsection.

This assumption means that two individuals, say i and j, who are not yet treated at date t have the same probability of leaving the state of interest at date t (and anytime before t). This should be true even if $t = 99$ days, $Z_i = 100$ days and $Z_j =$1 000 days i.e. the fact that i is going to be treated tomorrow does not make him more likely to leave today than j, who will be treated in more than two years. The NA assumption is violated if individuals receive information shocks before being treated and act on this new information. For instance, if i knows that he will be treated tomorrow and consequently stops searching for jobs, his hazard rate will differ from that of j. The main purpose of this paper is to test the NA assumption, using treatment notification dates as information shocks.

Before introducing notification dates, we briefly present the Timing of Events (ToE hereafter) framework which has become standard in the evaluation of dynamic treatments. The NA assumption coupled with an unconfoundability assumption are necessary to identify the model. If the latter assumption involves unobserved confounders, we also need to make a parametric assumption on the hazard rates of Z and Y in order to identify the distribution of unobserved heterogeneity. The ToE model typically assumes mixed-proportional hazard (MPH thereafter) rates:

$$\begin{aligned} \widetilde{h}_Z\left(t|X,\widetilde{V}\right) &= \widetilde{\lambda}_Z(t)\widetilde{\phi}_Z(X)\widetilde{V}_Z, \\ \widetilde{h}_Y\left(t|Z,X,\widetilde{V}\right) &= \lambda_Y(t)\phi_Y(X)\widetilde{V}_Y\left[\widetilde{\delta}_Z(t,Z,X)\right]^{\mathbf{1}\{Z<t\}}, \end{aligned} \tag{2}$$

where X is a vector of observed characteristics, $\widetilde{V} = \left(\widetilde{V}_Z, \widetilde{V}_Y\right)$ denotes unobserved confounders, $\widetilde{h}_Z\left(\cdot|X,\widetilde{V}\right)$ is the conditional hazard rate of Z and $\widetilde{h}_Y\left(\cdot|Z,X,\widetilde{V}\right)$ the hazard rate of Y, conditional on individual heterogeneity and on the treatment date. The distribution of $\widetilde{V}$, denoted as $\widetilde{G}(\cdot)$, is assumed to be independent of X. Using standard identification results from the competing risks literature (see Abbring and Van den Berg, 2003b) one can identify the $\widetilde{\lambda}$'s, the $\widetilde{\phi}$'s and $\widetilde{G}$ from data on $\min(Z,Y)$ and $I\{Y<Z\}$. Then, the variation of the hazard rate of Y around the treatment date identifies the treatment effect $\widetilde{\delta}$. Note that the dummy $\mathbf{1}\{Z<t\}$ in the hazard rate of Y ensures that the NA assumption is satisfied.

2.2 A duration model with notification dates

We here extend the standard evaluation model (2). Assuming that an individual enters the state of interest at date $t = 0$, we denote as P the duration until he receives some information about his future entry into treatment. We denote as $Z(p)$ the duration elapsed before the individual is treated, if he is notified at date $t = p$. Lastly, let $Y(z,p)$ be the time spent in the state of interest if the individual is treated at date $t = z$ and receives a notification at date $t = p$. The durations $Z(p)$ and $Y(z,p)$, $(z,p) \in (\mathbb{R}^{*+})^2$ are potential random duration variables. We assume that (P, Z, Y) is ruled by the following MPH model:

$$\begin{aligned}
h_P(t|X,V) &= \lambda_P(t)\phi_P(X)V_P, \\
h_Z(t|P,X,V) &= \lambda_Z(t)\phi_Z(X)V_Z\left[\gamma_P(t,P,X)\right]^{\mathbf{1}\{P<t\}}, \\
h_Y(t|Z,P,X,V) &= \lambda_Y(t)\phi_Y(X)V_Y \\
&\times \left[\delta_Z(t,Z,X)\right]^{\mathbf{1}\{Z\leq P\}\cdot\mathbf{1}\{Z<t\}}\left[\delta_P(t,P,X)\right]^{\mathbf{1}\{P<t\leq Z\}}\left[\delta_{PZ}(t,P,Z,X)\right]^{\mathbf{1}\{P<Z<t\}}.
\end{aligned} \tag{3}$$

where $h_A(\cdot|B)$ is the hazard rate of A conditionally on B, X is a vector of observed individual characteristics and $V = (V_P, V_Z, V_Y)$ is a vector of unobserved individual characteristics. We assume that V is a vector in $(\mathbb{R}^{*+})^3$, independent of X and with distribution G. It is important to mention that the functions λ_Z, λ_Y, ϕ_Z, ϕ_Z and G are not the same as those in the standard model (2). The set of functions $(\gamma_P, \delta_P, \delta_Z, \delta_{PZ})$ describes the effects of notification and treatment on the durations. We will comment on these effects in the next subsection. We add to model (3) a series of technical assumptions about continuity of the ϕ functions and about integrability of the λ, γ and δ functions (as well as cross products of these functions). We present these assumptions in Appendix A. From now on, any reference to model (3) will include this set of technical assumptions.[7]

We should mention that model (3) implies a specific notification procedure that might not be suitable for all applications. The main two features are: i) one could receive a treatment without having been previously notified (Z can be lower than P) and ii) the date of start of the treatment is still random once notification has been received (the distribution of Z is not degenerate if $Z > P$). These two characteristics of our model are introduced in prevision of our empirical application. One could also think of an alternative model in which notification necessarily comes before treatment and the date of treatment is deterministic conditionally on the notification date.

Model (3) satisfies two assumptions that are important for identification. First, unconfoundedness (CIA thereafter, for conditional independence assumption) states that within specified groups of individuals, treatment is assigned independently of potential outcomes. In model (3), we assume that $Y(z,p) \perp (Z,P)$ and $Z(p) \perp P$ conditionally on X and V.

Secondly, we need to adapt the NA assumption to our setting. The prescription P can have an effect on Z and Y while the actual treatment Z can have an effect only on Y.[8] Model

[7] The event-history model derived by Abbring (2008) would also be relevant to study transitions between our five states of interest (neither notified nor treated, notified and not treated, treated and not notified, notified and treated, and out). The modeling of the effects of notification and treatment would be slightly different though.

[8] For conciseness, we will refer to the effect of notification on (duration before) training as the effect of P on Z. Likewise, the effect of training on unemployment duration will be referred to as the effect of Z on Y, etc.

(3) implies that these "treatments" start having an effect on the "outcome" only from the date of their realization:

$$
\begin{aligned}
\Pr\left(Z(p)=t\right) &= \Pr\left(Z(p')=t\right), \qquad \forall t \le \min(p,p'), \\
\Pr\left(Y(z,p)=t\right) &= \Pr\left(Y(z,p')=t\right), \qquad \forall t \le \min(p,p') < z, \\
\Pr\left(Y(z,p)=t\right) &= \Pr\left(Y(z',p)=t\right), \qquad \forall t \le \min(z,z') < p, \\
\Pr\left(Y(z,p)=t\right) &= \Pr\left(Y(z',p)=t\right), \qquad \forall p < t \le \min(z,z'),
\end{aligned}
\tag{4}
$$

where we have dropped the conditioning on X and V for notational convenience. The first equality will allow us to identify the effect of P on Z. The second and third equalities pertain to the effect of P on Y and Z on Y respectively. The last equality will allow identification of the joint effect of P and Z on Y. Identification will be discussed in subsection 2.4.

To avoid confusion, we will not refer to (4) as no anticipation assumptions even though these equalities restrict the information available to individuals. For instance, the standard NA assumption discussed in section 2.1 states that no private information shocks is used by individuals before they start the treatment Z. Our model (3) allows for such a shock, through P, but we need to impose that notification is the only source of information prior to the treatment.

2.3 Three issues arising from notification

The main purpose of the present paper is to study whether notification shocks act as private signals that may violate the standard NA assumption. In addition to the study of anticipation of treatment by individuals, we want to know if the treatment effect depends on whether the individual has received a prescription or not. If this is the case then the standard framework of AVdB (2) no longer applies as the treatment effect δ depends on a time-varying unobserved variable (notification status). The next three paragraphs address specific issues related to notification and/or anticipation and point at potential biases in the estimation of the treatment effect when one overlooks, or has no information on, notification. The tests that are consequently suggested are based on the set of "treatment effect" functions $(\gamma_P, \delta_P, \delta_Z, \delta_{PZ})$.

Does notification have a direct effect on the outcome of interest? The main implication of the NA assumption is that before entering a treatment, individuals do not receive information shocks at random dates that will affect their propensity to leave the state of interest. However, if not yet treated individuals receive some information about their future participation in the treatment and act on this information, the NA assumption is violated. Note that the information does not need to be the actual date when an individual will be treated, it can just consist in some indication that the individual is more or less likely to receive the treatment at any date in the future. What is important is that individuals' information about their future treatment status can change over time, and that individuals can act on these information shocks.

We consider that notification can be seen as an information shock that could affect the individual's main duration outcome prior to treatment. It is clear from model (3) that notification has a direct effect on the outcome of interest if and only if $\delta_P(t, p, X) \neq 1$. One of our main empirical goals will thus be to test whether $\delta_P(t, p, X)$ is different from 1.

Does the treatment effect depend on notification? In theory, the model (2) allows for the treatment effect $\widetilde{\delta}_Z$ to depend on the treatment date, the time elapsed since treatment and individual observed characteristics X. AVdB also show that one can identify a treatment effect that depends on unobservable characteristics, provided these characteristics can only change at the treatment date. However, if the effect of the treatment differs whether individuals have received a prescription or not, and if the notification date is not available or overlooked, the treatment effect depends on an unobservable variable that changes through time, strictly before the treatment date. In that case, model (2) no longer allows for enough flexibility in the treatment effect. We can check empirically whether the treatment effect depends on the individual's notification status at the treatment date by testing the following equality $\delta_Z(t, z, X) = \delta_{PZ}(t, z, p, X)$.

Does notification have a direct effect on the treatment date? If individuals who receive notification are more likely to be treated, there is an unobserved heterogeneity term (the notification date P) that is unaccounted for in model (2). Therefore the MPH specification in (2) might be inconsistent with the true hazard rate of Z. We can check this by testing whether γ_P is different from 1. This issue is not directly related to anticipation behavior and is perhaps of a lesser methodological interest than the two effects discussed above. Yet, γ_P can be of considerable interest to policy makers as its sign and magnitude convey a lot of information on the efficiency of the treatment assignment process.

2.4 Identification

We always observe X and Y, although the latter can be censored by the sampling date.[9] However, we face the standard selectivity issue as we observe Z only for those who receive the treatment before leaving the state of interest, i.e. those who have $Z < Y$. If an individual

[9]This censoring affects few observations in our empirical application.

leaves before having been treated, we only know that $Z \geq Y$. Likewise, we observe P if and only if $P < \min(Z, Y)$. If an individual starts treatment or leaves the state of interest without having received notification, we only know that $P \geq \min(Z, Y)$. Formally, for any $(p, z, y) \in (\mathbb{R}^{*+})^3$ we can compute the four following probabilities:

$$\begin{aligned} Q_Y(y) &= \Pr\left(Y > y, Y < \min(\text{P,Z})\right), \\ Q_Z(z, y) &= \Pr\left(Y > y, Z > z, Z < \min(P, Y)\right), \\ Q_P(p, y) &= \Pr\left(Y > y, P > p, Z > Y > P\right), \\ Q_{PZ}(p, z, y) &= \Pr\left(Y > y, Z > z, P > p, Y > Z > P\right). \end{aligned} \tag{5}$$

These probabilities are conditional on X, we drop the conditioning for notational convenience. We denote $Q(p, z, y) = \big(Q_Y(y),\, Q_Z(z, y),\, Q_P(p, y),\, Q_{PZ}(p, z, y)\big)$ and will now refer to $Q = \left\{Q(p, z, y), (p, z, y) \in (\mathbb{R}^{*+})^3\right\}$ as "the data". We also define two subsets of data that will prove relevant for identification. The first one consists of the minimum of (P, Z, Y) and of an indicator telling which of these three processes is the shortest. Formally, we can write this subset as: $Q^0 = \left\{Q^0(p, z, y), (p, z, y) \in (\mathbb{R}^{*+})^3\right\}$ where $Q^0(p, z, y) = \big(Q_Y(y),\, Q_Z(z, -\infty),\, Q_P(p, -\infty),\, Q_{PZ}(p, -\infty, -\infty)\big)$. The subset Q^0 thus provides the duration up to the first event as well as the nature of this event (notification, treatment or exit from the state of interest). The second subset of interest consists of P (unless censored), the minimum of (Z, Y) and an indicator telling whether $Z < Y$ or not. Formally, we can write it as $Q^1 = \left\{Q^1(p, z, y), (p, z, y) \in (\mathbb{R}^{*+})^3\right\}$ where $Q^1(p, z, y) = \big(Q_Y(y),\, Q_Z(z, -\infty),\, Q_P(p, y),\, Q_{PZ}(p, z, -\infty)\big)$. The only difference between Q and Q^1 is that in the latter, Y is censored by Z i.e. we no longer follow individuals once they have received the treatment. We can now give three identification results. [10]

The first one is taken from Abbring and Van den Berg (2003b) and states that we can identify model (3), except the effect of notification or treatment, on the minimum of the three processes:

Proposition 1. [*Abbring and Van den Berg (2003b)*] The functions $(\lambda_P, \lambda_Z, \lambda_Y, \phi_P, \phi_Z, \phi_Y, G)$ from model (3) are identified from Q^0.

This proposition shows that the "competing risks" part of the model allows identification of the three hazard rates, except their link arising through notification and/or treatment. We still need to recover the effect of P on Y, of P on Z, of Z on Y if $Z \leq P$ and of Z on Y if $P < Z$. To this end, we need to use the remaining data and look at the change in the hazard rate of the outcomes around the date of notification and the date of treatment. We

[10] See Abbring (2008) for the identification of an event-history model in a similar vein.

can do this in two steps. First, we consider the subset of data Q^1 and focus on the effects of P on Z and on Y. We have the following result:

Proposition 2. The functions $(\lambda_P, \lambda_Z, \lambda_Y, \phi_P, \phi_Z, \phi_Y, G, \gamma_P, \delta_P)$ from model (3) are identified from Q^1.
Proof: See Appendix B.

Proposition 2 states that the variations of the hazard rates of Z and Y around the notification date are enough to identify the effect of notification. This result is important from an empirical perspective as it will allow us to estimate the effect of notification without making assumptions about the δ_Z and δ_{PZ} functions i.e. about the effect of Z on Y. Indeed, while in theory we can identify treatment effects as flexible functions, in practice we have to impose some structure on the treatment effects for the estimation to be feasible. Proposition 2 thus shows that an estimation of the effects of notification and thus, according to subsection 2.3, the test of the NA assumption do not rely on a given specification for the effects of Z on Y. The model identified from Proposition 2 will be referred to as the partial-information model.

While the test of the NA assumption is the main purpose of this paper, we also want to evaluate the effect of the treatment Z. To this end, we need to consider the whole data and identify all the functions from model (3). We thus need the last identification result:

Proposition 3. The model (3) is identified from the data Q.
Proof: See Appendix B.

As it was the case for the identification of γ_P and δ_P in Proposition 2, Proposition 3 states that we can identify the effects of Z, i.e. the functions δ_Z and δ_{PZ} by using the variation of the hazard rate of Y around Z when $Z \leq P$ and $P < Z$ respectively.

3 Empirical application: training programs for unemployed workers in France

3.1 Training programs and notification procedures in France

In this subsection we present the assignment process to training and the nature of the information shock individuals receive when they are notified. It appears that the institutional setting of the French training system is a source of variation for P and Z in our econometric model. We also give some insight on the content of training programs.

Notification: the nature of the "information shock". In France, entry into a training program may result from a proposal by the public employment service (Agence Nationale Pour l'Emploi, ANPE hereafter) or from the job seeker's own initiative. The PARE (Plan d'Aide au Retour a l'Emploi) reform implemented in 2001 improved individual counseling services. Since then, a meeting with an ANPE caseworker (typically 30 minutes long) is compulsory for all newly registered unemployed workers and recurs at least every 6 months. Depending on the individual's profile, the caseworker can schedule follow-up interviews between two compulsory meetings, and interviews can be requested at any moment by the unemployed workers themselves. Apart from a wide range of counseling measures, training programs may be proposed to job seekers during these interviews. This allows us to characterize notification in our econometric model. More precisely, notification is reported when an ANPE caseworker informs the job seeker that he should enter a training program and that he is to be put in relation with a (private or public) training provider.[11] In theory, notification should be given during, or shortly after, the second meeting with the caseworker (usually 6 months after registration). In practice, it can also occur during another meeting, or even by phone or (e-)mail. Hence notification can occur very early in the unemployment spell or much later, depending on the timing of interviews, but also on the discussions between the caseworker and the job seeker. In the framework of our econometric model, this can be seen as a source of variation in P, which will be supported by descriptive statistics in the next subsection.

From notification to training. When a job seeker is notified, he may not immediately, nor systematically, enter a training program. In theory, job seekers are free to accept or turn down any program they are proposed, but a refusal can lead to a cut in unemployment benefits. In practice, however, sanctions for refusing a training program are almost never taken.[12] Hence, notification implies no compulsory training action. This makes the French institutional setting very different from other systems where sanctions for a refusal of training are much more likely to occur.[13] Moreover, even if the job seeker is willing to be trained, finding a suitable program can take time. This is due to the lack of available training slots or to the time an individual needs to find a funding for his/her training program. Finally, despite recent reforms, the French training system remains complex[14] so notification is only

[11] It could be that the caseworker contacts the training provider on behalf of the job seeker or that he gives the job seeker the contact details of the training provider.

[12] Note that job seekers not eligible to unemployment benefits (roughly 50% of the stock) are not concerned by sanctions.

[13] See, e.g., the description of the Danish system in Rosholm and Svarer (2008).

[14] One of the main feature of the system is that it is run and funded by three different agents: the state, the social partners and the administrative regions. See, Crépon et al. (2007) for a more precise description

the first step in a long and possibly difficult procedure. In the next subsection we show that there is indeed a lot of variation in the duration between notification and treatment.

Notification and contents of training programs. Participation in a training program may or may not be preceded by notification from a caseworker. In the latter case, it could be that the job seeker found a training program on his own and then asked the caseworker to authorize it. There may thus be heterogeneity in the treatment effects with respect to who initiated training. It is not clear a priori how these two effects may differ. On the one hand, the job seeker has a better knowledge of his own skills, motivation and job experience but on the other hand, the caseworker has more information on the local labor market. For instance, since the PARE reform, ANPE caseworkers have access to detailed information on local labor demand and have been instructed to assign job seekers to training actions suited to the open vacancies (see Ferracci et al., 2010). Ideally, we would like to control for the actual content of training programs. This would allow us to interpret a difference between δ_Z and δ_{PZ} as the effect of the counseling effort of the caseworker. Unfortunately, this information is not available in our data so we shall work with a general definition of training programs, controlling for the notification status of the job seeker.[15]

3.2 Data and descriptive statistics

The data set. Our data come from the Fichier Historique Statistique (FHS thereafter), an exhaustive register of all unemployed spells recorded at the ANPE. We use data on 10% of individuals in the greater Paris region (Ile-de-France). We consider all unemployment spells starting in 2003 or 2004 and follow them up to their end or to the 1st of January 2008, which is the date when the data was extracted (very few spells last until then). For each spell we observe the starting and ending[16] dates (unless censored by the extraction date), an individual identifier and some characteristics of the job seeker (which we detail below). If an unemployment spell includes a period during which the individual follows a training

of the system.

[15] Additional data provided by the unemployment insurance agency (UNEDIC) make it possible to describe the content of training programs with some precision. Due to the lack of common identifiers, we cannot merge this additional data set with the one we use in this paper. This data set sorts training programs into four groups, according to the type of training: "*general*" (e.g. mathematics, economics, languages), "*personal*" (e.g. development of mental abilities, development of professional organization capacities), "*service oriented vocational skills*" (e.g. accounting, hotel business) and "*production oriented vocational skills*" (e.g carpentry, engineering). While the distribution across types is not uniform, the mass is not concentrated on a single type. For instance, out of the 593 126 programs that took place between 2005 and 2007, 17.9% were of the "*general*" type, 37.5% of the "*personal*" type, 29.9% were "*service oriented*" and 14.7% were "*production oriented*".

[16] An unemployment spell ends when the individual leaves the register of the ANPE which means either that he has found a job or that he has stopped looking for one.

program, we observe the dates when he enters and leaves this program. Importantly, we also know if and when the caseworker informs the job seeker of the action taken regarding his job search, and whether this involves taking steps towards a training program. As explained in the previous section, we consider that a job seeker has received notification of a future treatment when he is informed by the caseworker of the start of a procedure that should lead to a training program.

Description of the sample. We have N unemployment spells, each denoted by the index $i \in [1, N]$. For each spell i, we observe three dummies C_i^P, C_i^Z and C_i^Y indicating whether each duration of interest is censored or not. We observe the realized duration before notification P_i if $C_i^P = 0$ and we only know that this duration is longer than P_i if $C_i^P = 1$. We observe the realized duration before treatment Z_i if $C_i^Z = 0$ and we only know that this duration is longer than Z_i if $C_i^Z = 1$. We observe the realized unemployment duration Y_i if $C_i^Y = 0$ and we only know that this duration is longer than Y_i if $C_i^Y = 1$.

For each spell i, we observe some characteristics of the job seeker, which are denoted by the vector X_i. These characteristics are the following: $\mathbf{1}\{\text{male}\}$, age, age^2, exp, exp^2 (where exp is the experience in the occupation of the job searched), $\mathbf{1}\{\text{French}\}$, $\mathbf{1}\{\text{married}\}$, $\mathbf{1}\{\text{children}\}$, dummies for qualification (6 categories, the reference is "executive"), education (6 categories, the reference is "university degree") and department of residence (8 departments, the reference is Paris).[17] If t_{i0} is the date when the current spell starts, we also compute the individual's number of unemployment spells and time spent unemployed over the periods $[t_{i0} - 2 \text{ years}, t_{i0}[$ and $[t_{i0} - 5 \text{ years}, t_{i0} - 2 \text{ years}[$. Lastly, we use some precise information on the location of the unemployment agency to define an individual's local labor market and then to compute two indicators. Let y_{i0} be the year when spell i starts and let a_i be the location of the unemployment agency. The first indicator gives the proportion of unemployment spells in a_i which started during $y_{i0} - 1$ and saw training occur within one year. The second indicator gives the relative variation in the yearly inflow into unemployment for area a_i between years $y_{i0} - 1$ and y_{i0}.

Descriptive statistics. Our sample contains 159 599 unemployment spells, starting between the 1st of January 2003 and the 31st of December 2004. Only 2.61% of these spells are censored by the data extraction date (1st of January 2008). For 16 852 unemployment spells (10.6%), notification is received before the start of a training program or exit from unemployment.

[17]Departments are administrative areas smaller than regions and larger than municipalities, There are 95 departments in metropolitan France, and 8 in the region we study. Only Paris is both a municipality and a department.

Table 1 gives the proportion of spells containing a notification or a training period (or both) in the whole sample (first column) as well as in populations of a given gender, age or unemployment status during the two-year period preceding the current spell. We note that relatively few individuals are notified or trained (around 10%), that the proportion of treated is much greater among those who received a notification, and yet that many individuals enter a training program without having received prior notification from the caseworker. This latter feature can arise from heterogeneity in treatment as individuals might participate in training programs that are not provided by the unemployment agency. Unfortunately, the data do not allow us to observe the nature of the training programs. Yet, we will allow for this possibility by letting the effect of training vary between individuals who were notified and those who were not. Also, note that our modeling of the hazard rates for P and Z (see section 2.2, equation (3)), is consistent with the statistics shown in Table 1, in particular with $\Pr(Z < P) > 0$.

Table 1: Probabilities of receiving notification and/or training

	all sample	male	female	age ≤ 25	age ≥ 55	$X_{u2} > 0$	$X_{u2} = 0$
% notified	10.6	9.5	11.7	8.1	3.0	7.3	13.3
% treated	9.8	9.3	10.3	7.0	3.2	2.9	15.5
% treated if not notified	6.5	6.2	6.9	4.7	2.2	1.8	10.7
% treated if notified	37.4	38.7	36.4	33.8	38.3	16.2	47.1

Note: X_{u2} is the time spent unemployed in the two years preceding the start of the current spell.

Table 2 shows the average and a series of quantiles for the main durations of interest. We see that unemployment spells can be very long, with an average of almost one year ($E(Y) = 342$ days). Individuals who receive notification do so on average after 6 months, which is consistent with the interview process introduced by the PARE reform. Note though that there is variation in the date when notification is given. There is also a lot of variation in the starting date of training programs, with an average at around 8 months (247 days). For those who were given notification and actually started a training program, the interval between these two events is shorter than 3 months on average.

Unemployment duration seems to be affected by both training and notification. Indeed, we note that the average of Y is much smaller among individuals who did not participate in a training program than among those who did. Individuals who were neither notified nor treated also experience shorter unemployment spells than those who received notification (but no treatment). However, these numbers can be driven by observed and unobserved heterogeneity so we turn to our econometric model for a more detailed analysis of the effects of notification and training.

Table 2: Distribution of some durations of interest (in days)

	Mean	Q10	Q25	Q50	Q75	Q90
P if notified	181	9	28	107	250	454
Z if treated	247	46	98	196	350	526
Z if treated and not notified	236	35	88	182	336	515
Z if treated and notified	263	63	117	217	369	539
$Z - P$ if treated and notified	82	5	13	43	98	209
Y	342	29	68	211	495	865
Y if not notified and not treated	292	26	54	168	386	780
Y if notified and not treated	513	89	207	403	753	1120
$Y - P$ if notified and not treated	331	38	98	225	468	782
Y if treated	657	264	402	641	853	1088
Y if not treated	308	27	59	182	415	808
Y if treated and not notified	648	251	391	629	853	1085
Y if treated and notified	670	285	423	657	852	1090

3.3 Econometric specification and estimation procedure

The duration model. We use the K_P-quantiles of P conditionally on $C^P = 0$ as cut-off points for the piece-wise constant part of the hazard rate in (3). This introduces $K_P - 1$ parameters to estimate for λ_P, as we fix the probability on the first interval, λ_{P1} to be .001. We proceed similarly for λ_Z and λ_Y (except that we do not condition on $C^Y = 0$ for the latter), with $\lambda_{Z1} = .002$ and $\lambda_{Y1} = .004$. We set $K_P = K_Z = K_Y = 11$. The 30 parameters thus introduced are stacked in the vector Λ. The ϕ functions in (3) are specified as log-linear functions: $\phi_P(X) = \exp(X'\beta_P)$, $\phi_Z(X) = \exp(X'\beta_Z)$ and $\phi_Y(X) = \exp(X'\beta_Y)$.

Treatment effects. We allow the treatment effects to vary with time as follows:

$$\begin{aligned}
\gamma_P(t, P, X) &= \gamma_P^0 \cdot \mathbf{1}\{t \leq P + 92\} + \gamma_P^3 \cdot \mathbf{1}\{t > P + 92\}, \\
\delta_P(t, P, X) &= \delta_P^0 \cdot \mathbf{1}\{t \leq P + 183\} + \delta_P^6 \cdot \mathbf{1}\{t > P + 183\}, \\
\delta_Z(t, Z, X) &= \delta_Z^0 \cdot \mathbf{1}\{t \leq Z + 123\} + \delta_Z^4 \cdot \mathbf{1}\{Z + 123 < t \leq Z + 365\} + \delta_Z^{12} \cdot \mathbf{1}\{Z + 365 < t\}, \\
\delta_{PZ}(t, Z, P, X) &= \delta_{PZ}^0 \cdot \mathbf{1}\{t \leq Z + 123\} + \delta_{PZ}^4 \cdot \mathbf{1}\{Z + 123 < t \leq Z + 365\} + \delta_{PZ}^{12} \cdot \mathbf{1}\{Z + 365 < t\}.
\end{aligned} \tag{6}$$

Since durations are in days, equation (6) means that we allow for the effect of notification on treatment participation (resp. on exit from unemployment) to change after 3 months (resp. 6 months). Likewise, the effect of training on unemployment duration (whether it was preceded by notification of not) is allowed to change after 4 months and after a year. This latter feature aims at capturing a locking-in effect for training programs (i.e. workers spending less time on job search while being treated). The dates used in this specification of treatment effects are motivated by the descriptive statistics on durations in the previous subsection.

The distribution of unobserved heterogeneity. The distribution of unobserved heterogeneity G is assumed to have a discrete support with a given number R of mass points. More precisely $\Pr\left(V = \exp\left(v_r\right)\right) = p_r$, $\forall r \in [1, R]$, where $v_r \in \left(\mathbb{R}^{*+}\right)^3$. The probabilities are modeled as follows:

$$p_r = \frac{\exp(-\alpha_r)}{\sum\limits_{r=1}^{R}\exp(-\alpha_r)}, \quad \alpha_R = 0 \text{ and } \alpha_r \in \mathbb{R} \text{ if } r < R.$$

Note that this specification of unobserved heterogeneity is more flexible than the one usually encountered in empirical applications of the ToE approach (e.g. Van den Berg, Van der Klaauw and Van Ours, 2004, or Lalive et al., 2005) which assume that each component of the V vector can take a given number of values (often two) and then form groups as pairs (if there are two processes, triplets if there are three, etc...) of these values. The approach we retain here, and which we borrow from the statistical literature (see e.g. Aitkin, 1999), is more flexible and more suitable to cases like ours where there are more than two duration processes (as the number of parameters to estimate increases more slowly when we account for more groups). We should mention though than in theory, if we can increase the number of groups to infinity, the two methods are similar.[18]

The full-information likelihood. We stack $(\gamma_P^0, \gamma_P^3, \delta_P^0, \delta_P^6, \delta_Z^0, \delta_Z^4, \delta_Z^{12}, \delta_{PZ}^0, \delta_{PZ}^4, \delta_{PZ}^{12})$, Λ, β_P, β_Z, β_Y and $\{p_r, v_r\}_{r\in[1,R]}$ in the vector $\Theta(R)$. The contribution to the likelihood of spell i is given by:

$$\begin{aligned}
\ell\left(P_i, C_i^P, Z_i, C_i^Z, Y_i, C_i^Y | X_i, \Theta_R, R\right) &= \sum_{r=1}^{R} p_r \quad \left[h_P(P_i|X_i, v_r)\right]^{1-C_i^P} S_P(P_i|X_i, v_r) \\
&\quad \times \quad \left[h_Z(Z_i|X_i, v_r)\right]^{1-C_i^Z} S_Z(Z_i|X_i, v_r) \\
&\quad \times \quad \left[h_Y(Y_i|X_i, v_r)\right]^{1-C_i^Y} S_Y(Y_i|X_i, v_r),
\end{aligned} \qquad (7)$$

[18]The paper by Crépon et al. (2007) considers a factor-loading model for unobserved heterogeneity, which imposes more structure on the data but is equivalent to our method if the number of factors becomes large (which is hard to realize in practice).

where the hazard rates h_P, h_Z and h_Y are given by (3) and S_P, S_Z and S_Y denote the corresponding survival functions. For instance $S_P(t|X,V) = \exp\left[-\int_0^t h_P(u|X,V)du\right]$. We then define the full-information likelihood conditionally on the number of groups of unobserved heterogeneity:

$$L\left(\Theta_R, R\right) = \prod_{i=1}^{N} \ell\left(P_i, C_i^P, Z_i, C_i^Z, Y_i, C_i^Y | X_i, \Theta_R, R\right). \tag{8}$$

Our benchmark estimation sets the number of groups of unobserved heterogeneity R to 4, as further increases in R did not increase the likelihood.

The partial-information likelihood. As shown in Proposition 2, we can identify all the determinants of model (3) except the two functions δ_Z and δ_{PZ} using data in which Y is censored by Z. We can thus estimate the main effect of interest, δ_P, without specifying the δ_Z and δ_{PZ} functions. Our test of the NA assumption will not depend on a specific modeling of the effect of training. We can thus consider the subset of data in which $Y_i = Z_i$ and $C_i^Y = 1$ if $C_i^Z = 0$. The corresponding partial-information likelihood is similar to (8) except that (δ_Z, δ_{PZ}) are no longer in the vector of parameters. The maximization procedure is similar to the one outlined above.

4 Estimation results

The first subsection focuses on the effect of notification on unemployment duration and on participation in training programs. The second subsection shows the estimates of the treatment effects. It also draws comparisons with results obtained from standard ToE models, either overlooking notification (i.e. just looking at the effect of Z on Y) or assuming that the actual treatment starts when notification is given (i.e. looking at the effect of $\min(P, Z)$ on Y). For these two sections, we will consider the logarithms of the δ and γ parameters, so that a given effect is said to be negative (resp. null, resp. positive) when the corresponding log-parameter is smaller than (resp. equal to, resp. greater than) 0. The next three subsections present additional results on the role of time dependence, observed and unobserved heterogeneity in the three processes at play. These results are interesting as they shed light on the notification process which, as far as we know, has not yet been studied on such a large scale.

4.1 The effect of notification on exit from unemployment and treatment participation

The effect of notification on unemployment: a test of the no-anticipation assumption. Table 3 presents the estimates of $\ln \delta_P$, using the specification (6). First, note that the estimates from the partial- and full-information models are close, as they should be (cf. Proposition 2). Since the partial-information model puts no constraints on δ_Z and δ_{PZ}, we can thus conclude that the specification used for these two effects in the full-information model does not impact the estimation of δ_P.

The main result of this paper is that notification significantly decreases the unemployment hazard rate. We can see in Table 3 that the effect of notification on h_Y is significantly negative, at -.4, in the first six months and not significantly different from zero thereafter. Therefore, in our data, the assumption that workers do not anticipate training is violated. The main consequence from a methodological point of view is that we can no longer interpret the treatment parameter $\widetilde{\delta}_Z$ in the standard ToE model (2) as a causal treatment effect. Indeed, the distribution of the counterfactual duration $E(Y|Y < Z)$ is not unique, it depends on an unobserved, time-varying, characteristic: the date of notification P.

Let us now discuss the economic implications of the estimates shown in Table 3. Since $\ln \delta_P$ remains significantly below 0 during the first semester following notification, our results rule out a "threat effect" of notification (see, e.g., Black et al., 2003). If workers who are notified then face sanctions for not participating in a training program, they might leave unemployment for a job that they would not have accepted otherwise. This job may indeed offer a better alternative to entering a training program or facing sanctions for not doing so. We do not find such an effect in our data as workers are less likely to leave unemployment once they receive notification of a future treatment. There are two potential causes. First, workers could actually want to participate in a training program and thus stop their job search in the weeks prior to the beginning of the program, leading to a decrease in the arrival rate of job offers. Second, the notification can increase a worker's reservation value as it gives him another alternative to compare job offers with. Since, in practice, sanctions against those shunning away from training are almost never implemented, this training opportunity only makes workers more selective when considering job offers. We cannot assess which of these two effects drives our results. Yet, Table 3 clearly indicates that the NA assumption does not hold because workers who receive notification prior to training experience a significant decrease in their probability to leave unemployment.

Table 3: Effects of notification on unemployment duration ($\ln \delta_P$)

	if $t \leq P + 183$	*if* $t > P + 183$
partial-info model	$-.412$ (.03)	$-.022$ (.037)
full-info model	$-.402$ (.02)	$-.0014$ (.029)

Note that the decrease in the hazard rate only takes place in the first six months following notification, after which $\ln \delta_P$ goes from -.4 to 0. As Table 2 showed, 75% (resp. 90%) of individuals who are both notified and treated enter a training program within 92 days (resp. 209 days) following notification. We can thus expect notification to have a small effect on the unemployment hazard rate after six months. Table 3 confirms this intuition as δ_P is no longer significantly different from 1. Notification thus seems to "lock" workers in unemployment only on the short term.

The effect of notification on treatment participation. Table 4 shows the estimates of $\ln \gamma_P$ from both the partial- and full-information models. As it was the case for δ_P, the estimates are very close so our specification of the effects of training in the full-information model does not affect the estimation of γ_P. The main result from Table 4 is that notification has a huge effect on the probability to enter a training program. This was expected given the descriptive statistics in Table 1, which reflected the setting of training programs in France. If a job seeker is offered a training program, he will be notified by the caseworker before entering the program. The reasons why there is not a one-to-one relation between notification and treatment status are twofold. First, workers can leave unemployment between notification and training. This might be due not only to a "threat-effect" of training but also to some inefficiencies in the assignment to treatment as workers may have to wait before a suitable training position opens (see Fleuret, 2006). Second, workers can find a training program which is not preceded by a notification from the caseworker. Still, the results from Table 4 show that the probability to enter a training program is much higher once an individual has received a notification. We note that the hazard rate of Z increases dramatically during the three months following notification and then decreases but remains at a level much higher than before notification.

Table 4: Effects of notification on treatment participation ($\ln \gamma_P$)

	if $t \leq P + 92$	*if* $t > P + 92$
partial-info model	4.49 (.11)	3.47 (.11)
full-info model	4.45 (.10)	3.26 (.11)

As we discussed in section 2.2, while γ_P might be of interest to policy makers who want to assess the efficiency of the assignment process, it is less important than the δ's parameters from a methodological point of view. Still, if one uses the standard ToE model (2) with no information on notification, the jump in the hazard rate of Z shown in Table 4 at the time of notification will not be accounted for. There will thus be a bias due to some time-dependent unobserved heterogeneity. In other words, even if the no-anticipation held (which is not the case in our data), the fact that $\ln \gamma_P$ is significantly different from 0 would create a bias in the estimation of the treatment effect.

4.2 The effect of training programs on unemployment duration, with and without notification

Table 5 shows our estimates for the effects of training on exit from unemployment. The first row corresponds to training programs that were not preceded by notification. We see that the effect on exit from unemployment is negative in the first four months, slightly positive in the eight following months and increases further after a year. These results are consistent with the already existing evidence (see, e.g., Lechner et al., 2004) showing that training programs have a strong locking-in effect followed by a null or slight increase in the probability to leave unemployment. This increase appears mostly a year after entry into a training program. Since training programs not preceded by notification start on average eight months after the beginning of an unemployment spell (see Table 2), these programs have a small impact on unemployment duration and none (perhaps even a positive one) on the probability of becoming long-term unemployed.[19]

[19] We should mention that while training programs seem to lenghthen unemployment spells, it has also been found (see Crépon et al., 2007) that they have a positive impact on the length of the subsequent employment spell.

Table 5: Treatment effects without ($\ln \delta_Z$) and with ($\ln \delta_{PZ}$) notification

	if $t \leq Z + 123$	*if* $t \in]Z+123, Z+365]$	*if* $t > Z + 365$
$\ln \delta_Z$	$-.775$ (.040)	.125 (.033)	.372 (.037)
$\ln \delta_{PZ}$	-1.17 (.041)	$-.129$ (.025)	.148 (.028)

Looking at the second row of Table 5, we note that the results are different for training programs with a notification from the caseworker. The main difference is that δ_{PZ} is always smaller than δ_Z. The locking-in effect during the first four months is stronger (-1.17 against -.775) and the effect between four and eight months is small but now significantly negative (-.129 against .125). The effect after a year is positive but smaller than δ_Z. The differences between δ_Z and δ_{PZ} may arise from different sources. First, it could be that the training programs are different with respect to their content or their intensity (half-time/full-time).[20] We cannot investigate further this issue for lack of data on the content of programs or of a more flexible specification. The treatment effects may also differ if notification triggers a behavioral reponse from the job seeker when he starts the training program.[21] In any case, the estimates in Table 5 show that notification is a relevant feature for the evaluation of training programs as the treatment effect depends significantly (and negatively) on whether an individual received notification.

We now compare our estimates of the treatment effect with those from the standard ToE model. The first row of Table 6 shows the estimate of $\ln \widetilde{\delta}_Z$ obtained using model (2) with a specification similar to that retained for our benchmark model (3). Note that with model (2), there is a single treatment effect, denoted as $\widetilde{\delta}_Z$, while model (3) allows for two treatment effects: with and without notification. Looking at the point estimates, we note that $\widetilde{\delta}_Z$ is between δ_Z and δ_{PZ} when $t - Z \leq 123$ days. Then $\ln \widetilde{\delta}_Z$ is above $\ln \delta_Z$ (and thus $\ln \delta_{PZ}$) between the 4$^{\text{th}}$ and 12$^{\text{th}}$ months that follow Z, and even more so after that (.503 against .372). Hence the treatment effect tends to be overestimated after 4 months, and especially after a year, when overlooking notification. While the two models lead to qualitatively similar

[20]While we do not observe the content of training programs, we know when each program starts and ends so we can compute its duration. The distribution of these durations barely changes when looking at programs with or without notification. The average and median are respectively 136 and 102 days with notification and 132 and 95 days without notification.

[21]The differences between δ_Z and δ_{PZ} may also be due to a misspecification of the model. Individuals who go directly into training (without receiving notification) may differ from those who receive notification in a way that is not captured by V in our MPH model.

estimates, we should not conclude that one can ignore notification when evaluating a given treatment. First, the violation of the no-anticipation assumption means that the estimates from the ToE model cannot be interpreted as a treatment effect because the counter-factual is ill-defined. In particular, the estimates shown in Table 6 are based on a model that is rejected by the data. Second, Table 5 shows that there may be heterogeneity between training programs so the estimates from the ToE model leads to some average effect and overlooks the selection between each types of program. Lastly, the long-run effect of training programs are overestimated when using model (2).

Table 6: Treatment effects using the standard ToE model $\left(\ln \widetilde{\delta}_Z\right)$

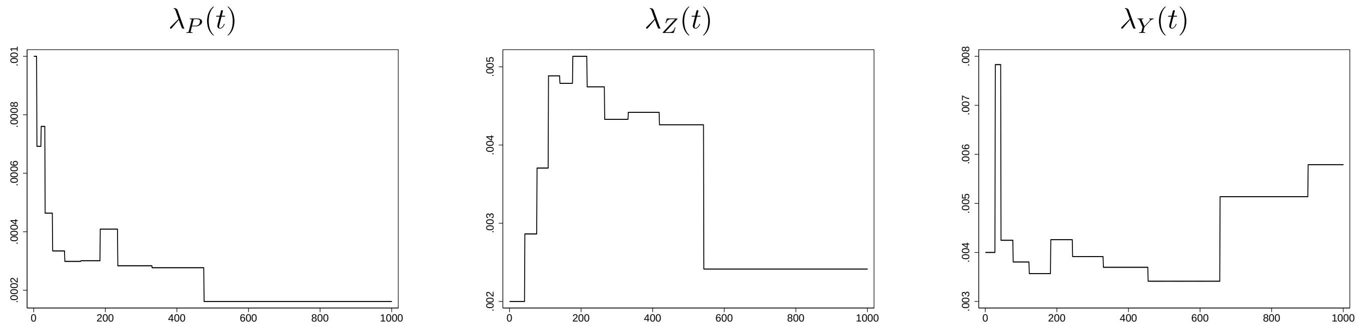
	if $t \leq Z+123$	*if* $t \in]Z+123, Z+365]$	*if* $t > Z+365$
model (2)	$-.824$ (.031)	.155 (.028)	.503 (.035)
model (2) with $Z = \min(P, Z)$	$-.557$ (.02)	.025 (.02)	.307 (.03)

The second row of Table 6 shows the estimate of the standard ToE model (2) when setting the treatment date to the minimum of the notification date P and the training date Z. This specification is sometimes suggested (for instance in AVdB) as a way to define the treatment in the presence of information shocks. We note that the locking-in effect during the first four months is weaker than in Table 5 (i.e. $\ln \widetilde{\delta}_Z$ is closer to 0). Also, the effect between 4 and 12 months is now insignificant. However, we note that this specification pushes the bias in the long-term treatment effect downwards (.307 against .503). Overall, it is difficult to interpret this treatment effect, as it is a compound of two effects (notification and training), and it shows no straightforward relation with the actual training effects estimated in Table 5.

4.3 Time dependence

We now look at the estimates of the λ functions, which, together with the treatment parameters, are the only source of time dependence in model (3). Figure 1 shows these functions for each duration of interest. When looking at these graphs, remember that we had to set the probability on the first interval for each process so these results are only qualitative. Remember also that the cut-off points have been set in order to match the deciles of P (conditionally on receiving notification), Z (conditionally on being treated) and Y.

Figure 1: Time-dependent components of hazard rates (t in days)



We note that workers are more likely to receive notification upon entering unemployment and also after about 200 days. These results are consistent with the timing of interviews as job seekers have to meet with a caseworker at the beginning of the unemployment spell and about six months later. The piecewise constant component of h_Z is low during the first weeks but increases steadily to reach a maximum after about 200 days, which is the period with the closest monitoring of job seekers. After this peak, λ_Z stays almost constant up to 550 days and then slumps so that almost no one is treated after 18 months of unemployment. Lastly, the hazard rate out of unemployment also depends on time as λ_Y shows a peak after about a month of unemployment. This non-stationarity in the probability to leave unemployment arises from worker reallocation between jobs through very short unemployment spells (see Fougère, 2000). After another albeit much smaller peak at 200 days, λ_Y shows a steady decline until $t \approx 650$ days where it jumps to much higher values. This could reflect the end of unemployment benefits (usually after 23 months of unemployment).

4.4 Observed heterogeneity

We now present the estimates for the effects of observed characteristics on the three hazard rates. Table 7 shows the estimates of the β parameters. Since we assumed a log-linear specification for the ϕ functions, a given characteristics is said to have a positive (resp. null, resp. negative) effect on the hazard rate when the corresponding parameter is positive (resp. null, resp. negative).

Table 7: The effect of observed characteristics on the three hazard rates

	β_P	β_Z	β_Y		β_P	β_Z	β_Y
$\mathbf{1}\{$male$\}$	$-.16^*$	.025	$.027^*$	$\mathbf{1}\{$blue col$\}$	$.25^*$	$-.21^*$	$.096^*$
age	$.2^*$	$.25^*$	$-.3^*$	$\mathbf{1}\{$white col unsk$\}$	$.39^*$	$-.33^*$	$.071^*$
age^2	$-.28^*$	$-.3^*$	$.21^*$	$\mathbf{1}\{$white col sk$\}$	$.26^*$	$-.14^*$	.0027
exp	$-.31^*$	$-.11^*$	$-.11^*$	$\mathbf{1}\{$technical$\}$	$.28^*$	.0017	$-.061^*$
exp^2	$.18^*$	.00014	$.098^*$	$\mathbf{1}\{$supervisor$\}$	$.22^*$	$-.021$	.011
$\mathbf{1}\{$French$\}$	$-.16^*$	$.33^*$	$-.16^*$	$\mathbf{1}\{$jr hs drop out$\}$	$-.069^*$	$-.56^*$	$.13^*$
$\mathbf{1}\{$children$\}$	.028	$.1^*$	$-.031^*$	$\mathbf{1}\{$jr hs degree$\}$	$.1^*$	$.096^*$	.02
$\mathbf{1}\{$married$\}$	$-.0047$	$-.12^*$	.013	$\mathbf{1}\{$hs drop out$\}$	$.18^*$	$-.011$	$.031^*$
$\mathbf{1}\{$dep. 77$\}$	$.14^*$	$.38^*$	$-.056^*$	$\mathbf{1}\{$hs degree$\}$	$.18^*$	.02	$-.029^*$
$\mathbf{1}\{$dep. 78$\}$	$-.066^*$	$.19^*$	$-.016$	$\mathbf{1}\{$univ drop out$\}$	$.24^*$	$.14^*$	$-.043^*$
$\mathbf{1}\{$dep. 91$\}$	$-.036$	$.42^*$	$.023^*$	% treated last year	-3.1^*	1^*	.14
$\mathbf{1}\{$dep. 92$\}$	$.25^*$	$-.0096$	.011	growth of u. inflow	$-.03$	$-.086$	$-.38^*$
$\mathbf{1}\{$dep. 93$\}$	$-.22^*$	$.17^*$	$-.087^*$	# u spells in $[t_0-2,t_0]$	$-.071^*$	$-.69^*$	$.24^*$
$\mathbf{1}\{$dep. 94$\}$	.015	$.27^*$	$-.0079$	time u in $[t_0-2,t_0]$	$-.18^*$	$-.34^*$	$-.037^*$
$\mathbf{1}\{$dep. 95$\}$	$.34^*$	.068	$-.094^*$	# u spells in $[t_0-5,t_0-2]$	$-.038^*$	$-.34^*$	$.076^*$
				time u in $[t_0-5,t_0-2]$	$-.13^*$	$-.38^*$	$-.11^*$

exp: experience, col: collar, sk: skilled, unsk: unskilled, jr: junior, hs: high school, univ: university and u: unemployment. A star means that the estimate is significant at the 5% level.

We note that the three hazard rates depend on both age and experience. Married job seekers are less likely to be treated whereas those who have children enter training programs more quickly. None of these two characteristics affects the notification process. Looking at the department dummies (dep. 77 to dep. 95), we see that the department 93, which has a high unemployment rate, shows a slower notification process but a faster assignment to training programs than Paris (the reference). On the contrary, in the department 92, which includes the business center of La Défense and some of the wealthiest Parisian suburbs, individuals are more likely to receive notification but face the same probability of being treated or of leaving unemployment than Parisians. Looking at workers' qualifications, we see that executives (the reference) receive fewer notifications but are more likely to enter a training program than other workers. This result confirms those of a recent field study (Fleuret, 2006) on the assignment process to training. At the local unemployment agency level, we note that agencies with a higher proportion of treated during the previous year will also have a higher proportion of treated in the present year but a lower notification rate. Lastly, individuals who spent more time unemployed or experienced more unemployment spells over the last seven years are less likely to receive notification or to enter a training program.

4.5 Unobserved heterogeneity

Table 8 shows the estimated probabilities for each group together with the (log-)effects of unobserved heterogeneity on the three hazard rates. We note that group 4 has a probability close to zero so the population is mainly split into three groups. Group 2 shows the lowest propensity to be treated or to leave unemployment and a relatively low propensity to receive notification. It is also much smaller than the other two groups, 1 and 3. It may reflect workers who are the most excluded from the labor market as they almost never participate in a training program and are much more likely to experience long-term unemployment.

We now focus on groups 1 and 3 which represent more than 90% of the population. Group 1 shows a relatively lower (resp. higher) propensity to be notified (resp. to enter a training program). Group 3, the largest group in terms of probability, contains individuals with the highest propensity to be notified but a lower propensity to enter a training program than those in group 1. This could be interpreted as efforts from the caseworkers to offer training programs to workers with a smaller propensity to look for such programs on their own. We also note that workers in group 3 are more likely to leave unemployment quickly but much less prone to participate in a training program than workers in group 1. We may thus expect the estimates of the treatment effects to be affected by the presence of unobserved heterogeneity. More precisely, if we do not account for such heterogeneity, we may underestimate the effect of notification on training as well as the direct effect of training on exit from unemployment. However, overlooking unobserved heterogeneity may lead to an upwards bias of the effect of notification on exit from unemployment. Since this latter effect was found to be negative (cf. Table 3) this bias may lead to accepting the NA assumption ($\ln \delta_P = 0$) while it is in fact rejected by the data.

Table 8: The distribution of unobserved heterogeneity ($R = 4$)

	p_r	$\ln V_P$	$\ln V_Z$	$\ln V_Y$
group 1	.309 (.03)	$-.959$ (.11)	-2.04 (.1)	$-.467$ (.04)
group 2	.079 (.01)	$-.547$ (.09)	-14.57 (2.56)	-1.61 (.04)
group 3	.605 (.03)	.529 (.06)	-5.05 (.12)	$-.135$ (.02)
group 4	.007 (.0009)	$-.053$ (.25)	1.92 (.16)	-1.11 (.06)

To assess further the importance of unobserved heterogeneity and illustrate the conclusions from Table 8, we present in Table 9 a comparison between two series of estimates of (logarithms of) the main parameters: δ_P, γ_P, δ_Z and δ_{PZ}. For each parameter, the number on the left is the estimate from model (3) with no unobserved heterogeneity i.e. $R = 0$ while the number on the right is the estimate from the same model with our benchmark specification of $R = 4$ groups of unobserved heterogeneity (which were the results shown in Tables 3-5).

First, we note that the predictions from Table 8 are true i.e. the "homogenous" model ($R = 0$) overestimates δ_P and underestimates γ_P and (except on the long term) δ_Z. The bias on δ_{PZ} (training preceded by notification) is more difficult to interpret because it involves the three duration processes.

A closer look at Table 9 reveals that the model with no unobserved heterogeneity leads to qualitatively different conclusions for two parameters. Indeed, it predicts that notification still has an effect on unemployment after 6 months ($\ln \delta_P = .127$ with a standard error of .01) whereas it has no effect in the model where we set $R = 4$. Also, the homogenous model points at a significantly positive effect of training preceded by notification on unemployment after 4 months ($\ln \delta_{PZ} = .147$ with a standard error of .02), whereas the inclusion of unobserved heterogeneity reveals that between the 4^{th} and 12^{th} months that follow training, workers are still "locked in" unemployment ($\ln \delta_{PZ}$ significantly negative at -.129).

In addition to these qualitatively biased conclusions, using the homogenous model leads to a potentially large bias in the point estimates of the treatment effects. For instance, the (logarithm of the) effect of notification on training after 3 months (γ_P) increases by a factor 3 when going from $R = 0$ to $R = 4$. The largest relative difference between point estimates lies in the long-term effect of training preceded by notification on exit from unemployment. The last row of Table 9 shows that $\ln \delta_{PZ}$ is 4 times smaller when introducing 4 groups of unobserved heterogeneity. The fact that the largest biases are found for the parameter δ_{PZ} is not surprising as the related treatment (notification followed by training) and outcome (exit from unemployment) involve the three duration processes.

Table 9: Estimates without ($R = 0$) and with ($R = 4$) unobserved heterogeneity

	$R=0$		$R=4$	
$\ln \delta_P$:	$-.252$ (.02)	if $t \leq P + 183$	$-.402$ (.02)	if $t \leq P + 183$
	.127 (.01)	if $t > P + 183$	$-.0014$ (.03)	if $t > P + 183$
$\ln \gamma_P$:	2.62 (.02)	if $t \leq P + 92$	4.45 (.10)	if $t \leq P + 92$
	1.09 (.03)	if $t > P + 92$	3.26 (.11)	if $t > P + 92$
$\ln \delta_Z$:	$-.877$ (.03)	if $t \leq Z + 123$	$-.775$ (.040)	if $t \leq Z + 123$
	.101 (.02)	if $Z + 365 \geq t > Z + 123$	.125 (.033)	if $Z + 365 \geq t > Z + 123$
	.495 (.02)	if $t > Z + 365$	.372 (.037)	if $t > Z + 365$
$\ln \delta_{PZ}$:	$-.984$ (.04)	if $t \leq Z + 123$	-1.17 (.041)	if $t \leq Z + 123$
	.147 (.02)	if $Z + 365 \geq t > Z + 123$	$-.129$ (.025)	if $Z + 365 \geq t > Z + 123$
	.591 (.02)	if $t > Z + 365$	.148 (.028)	if $t > Z + 365$

5 Conclusion

The methodological contribution of this paper is to extend the standard ToE model to allow for information shocks that may precede participation to treatment, and to present the assumptions under which the identification of each effect of interest is achieved. Our dynamic setting is thus more general than the previous studies conducted on anticipation effects of training programs and could be taken to other empirical applications. An important result is that the assumption that individuals do not anticipate future treatments, which numerous evaluations of active labor market policies (ALMPs) rely on, is violated in our data. Estimations based on the ToE approach that do not take the notification process into account may thus be biased. This is a case for restricting the validity conditions of the standard ToE approach when data on notification are not available. Using a partial-information model and the results we derived, we run a conclusive robustness check that shows that our results on individuals' anticipation behavior do not depend on our specification of the treatment effects.

As for policy implications, our paper shows that notification of future training programs actually decreases the exit rate from unemployment. This result goes against several other evaluations, which provide evidence of some "threat effect" of ALMPs. Yet, this fact has to be interpreted in the light of the French institutional setting, in which training is not mandatory. In most cases, French job seekers rarely face sanctions if they do not participate in a training program, so it is hardly surprising that notification tends to lengthen their unemployment duration. On the other hand, French policy makers could consider imposing mandatory training measures from a predetermined date in the unemployment spell. This could help drawing the line between those job seekers whose search effort is high, but unfruitful, and those with a higher preference for leisure, who will presumably exit unemployment on their own at the approach of entry into training. More generally, future research on ALMPs should use data on the assignment process to study the anticipation of treatments and thus offer a more detailed analysis of the effects of training programs.

A natural extension to the present analysis would consist in taking a closer look at the anticipation behavior of unemployed workers and try to understand their response to information shocks. Such an analysis could build on structural job search models (see for instance the recent contribution by Van den Berg et al., 2009) and allow the workers' reservation values to depend on the opportunity to enter a training program. One interpretation of our result that goes against a threat effect of notification could indeed be that, due to the non-enforcing nature of training offers, notification just presents job seekers with a new alternative that will increase their reservation value and thus lenghten their unemployment spell. It would be interesting to take this interpretation to the data using a structural model.

References

Abbring, J. (2008): "The Event-History Approach to Program Evaluation," in D. Millimet, J. Smith, and E. Vytlacil, eds, Advances in Econometrics, Volume 21: Modeling and Evaluating Treatment Effects in Econometrics, Elsevier Science, Oxford.

Abbring, J., and G.J. van den Berg (2003a): "The Non-Parametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71, 1491-517.

Abbring, J., and G.J. van den Berg (2003b): "The Identifiability of the Mixed Proportional Hazards Competing Risks Model," *Journal of the Royal Statistical Society, Series B*, 65, 701-710.

Abbring, J., and J. Heckman (2008): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in: Heckman, J., and E. Leamer, eds., *Handbook of Econometrics*, vol. 6B, ch. 72. Amsterdam: Elsevier Science, 5145-303.

Aitkin, M. (1999): "A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models," *Biometrics*, 55(1), 117-28.

Arni, P., Lalive, R., and J. van Ours (2009): "How Effective are Unemployment Benefit Sanctions? Looking Beyond Unemployment Exit," *IZA Discussion Paper*, n° 4509.

Black, D., Smith, J., Berger, M., and B. Noel (2003): "Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System," *American Economic Review*, 93(4), 1313-27.

Cockx, B., Dejemeppe, M. (2007): "Is the Notification of Monitoring a Threat to the Unemployed? A Regression Discontinuity Approach," *IZA Discussion Paper*, n° 2854.

Crépon, B., Ferracci, M. and D. Fougère (2007): "Training the Unemployed in France: How Does It Affect Unemployment Duration and Recurrence?" *IZA working paper*, 3215.

Crépon, B., Jolivet, G., Ferracci, M., and G.J. van den Berg (2009): "Active Labor Market Policy Effects in A Dynamic Setting", *Journal of the European Economic Association*, 7(2-3), 595-605.

De Giorgi, G. (2005): "Long-Term Effects of a Mandatory Multistage Program: The New Deal for Young People in the UK," Working paper, IFS, London.

Eberwein, Ham and LaLonde (1997): "The Impact of Being Offered and receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data," *Review of Economic Studies*, 64, 655-682.

Ferracci, M., Jolivet, G. and G.J. van den Berg (2010): "Treatment Evaluation in the Case of Interactions Within Markets," *IZA discussion paper*, 4700.

Fleuret, A. (2006): "L'Orientation des Chômeurs Vers la Formation de 2002 à 2004," *DARES Premières synthèses*, 29(2).

Fougère D. (2000): "La durée du chômage en France", in *Réduction du chômage : les réussites en Europe, Rapport du Conseil d'Analyse Economique*, 23. La Documentation Française, Paris, 239-59.

Fredriksson and Johansson (2008): "Dynamic Treatment Evaluation," *Journal of Business and Economic Statistics*, 26, 435-445.

Geerdsen, L. (2006): "Is There a Threat Effect of Labour Market Programs? A Study of ALMP in the Danish UI System," *Economic Journal*, 116, 738-50.

Heckman, J., Lalonde, R. and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in: Ashenfelter, O., and D. Card, eds., *Handbook of Labor Economics*, vol. 3A, ch. 31. North-Holland, New York, 1865-2097.

Heckman, J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 341-96.

Lalive, R., Van Ours, J., and J. Zweimüller (2005): "The Effect of Benefit Sanctions on the Duration of Unemployment," *Journal of the European Economic Association*, 3(6), 1386-417.

Lechner, M., Miquel, R. and C. Wunsch (2004): "Long-Run Effects of Public Sector Sponsored Training in West Germany," *Journal of the European Economic Association*, forthcoming.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles," translated (with discussion) in *Statistical Science*, 5(4), 465-80.

Richardson, K. and G.J. van den Berg (2006): "Swedish Labor Market Training and the Duration of Unemployment," *IZA discussion paper*, 2314.

Robins, J. (1997): "Causal Inference From Complex Longitudinal Data," *in* Berkane, M., ed., *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics.* Springer-Verlag, New York, 69-117.

Rosholm, M., and M. Svarer (2008): "Estimating the Threat Effect of Active Labour Market Programmes," *Scandinavian Journal of Economics*, 110(2), 385-401.

Rubin, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

Sianesi, B. (2004): "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," *Review of Economics and Statistics*, 86, 133-55.

Van den Berg, G.J., Bergemann, A. and M. Caliendo (2009): "The Effect of Active Labor Market Programs on Not-Yet Treated Unemployed Individuals," *Journal of the European Economic Association*, 7(2-3), 606-16.

Van den Berg, G.J., Bergemann, A. and M. Caliendo (2010): "The Effect of Training and Workfare on Not-Yet Treated Unemployed Individuals," Working paper, University of Mannheim.

Van den Berg, G.J., Bozio, A., and M. Costa Dias (2010): "Policy Discontinuity and Duration Outcomes," Working paper, University of Mannheim.

Van den Berg, G.J., Van der Klaauw, B. and J. Van Ours (2004): "Punitive Sanctions and the Transition Rate from Welfare to Work," *Journal of Labor Economics*, 22(1), 211-41.

APPENDIX

A Technical assumptions for the benchmark model.

Our benchmark model (3) extends the AVdB approach to the case of three processes. We here adapt the technical regularity assumptions of AVdB to our model in order to derive our identification results. These assumptions are:

- the three functions, ϕ_P, ϕ_Z and ϕ_Y are continuous from the space of X to $\mathbb{R}^{*+}$ and all equal to 1 for some value x^*,

- the three functions λ_P, λ_Z and λ_Y (all going from $\mathbb{R}^+$ to $\mathbb{R}^{+*}$) have finite integrals $\Lambda_P(t) = \int_0^t \lambda_P(\tau)d\tau$, $\Lambda_Z(t) = \int_0^t \lambda_Z(\tau)d\tau$ and $\Lambda_Y(t) = \int_0^t \lambda_Y(\tau)d\tau$. Also, there is a given $t^* \in \mathbb{R}^+$ such that $\Lambda_P(t^*) = \Lambda_Z P(t^*) = \Lambda_Y(t^*) = 1$.

- G is such that $\Pr\left(V \in (\mathbb{R}^{+*})^3\right) > 0$

- the γ and δ functions are such that the following functions exist:

$$\Delta_P(t,p,x) = \int_p^t \delta_P(t,p,x)dt, \qquad \Omega_P(t,p,x) = \int_p^t \lambda_P(t)\delta_P(t,p,x)dt,$$

$$\Gamma_P(t,p,x) = \int_p^t \gamma_P(t,p,x)dt, \qquad \Psi_P(t,p,x) = \int_p^t \lambda_Z(t)\gamma_P(t,p,x)dt,$$

$$\Delta_Z(t,z,x) = \int_z^t \delta_Z(t,z,x)dt, \qquad \Omega_Z(t,z,x) = \int_z^t \lambda_Y(t)\delta_Z(t,z,x)dt,$$

$$\Delta_{PZ}(t,p,z,x) = \int_z^t \delta_{PZ}(t,p,z,x)dt, \qquad \Omega_{PZ}(t,p,z,x) = \int_z^t \lambda_Y(t)\delta_{PZ}(t,p,z,x)dt,$$

and Δ_P, Ω_P, Γ_P and Ψ_P are continuous for all $p \in \mathbb{R}^+, t > p, x$, Δ_Z and Ω_Z are continuous for all $z \in \mathbb{R}^+, t > z, x$ and Δ_{PZ} and Ω_{PZ} are continuous for all $p \in \mathbb{R}^+, z > p, t > z, x$.

B Proofs of Propositions 2 and 3.

We first introduce the trivariate Laplace transform of G:

$$\mathcal{L}_G(u_1,u_2,u_3) = \int_0^{+\infty}\int_0^{+\infty}\int_0^{+\infty} \exp(-u_1 v_P - u_2 v_Z - u_3 v_Y)dG(v_P,v_Z,v_Y),$$

as well as the marginal derivatives: $\mathcal{L}_G^{(P)} = \partial\mathcal{L}_G/\partial u_1$, $\mathcal{L}_G^{(Z)} = \partial\mathcal{L}_G/\partial u_2$ and $\mathcal{L}_G^{(PZ)} = \partial^2\mathcal{L}_G/\partial u_1\partial u_2$. Note also that the following will use notations introduced in section A of the Appendix. Lastly, we now write the conditioning on x in the notations of the Q functions from (5).

Proof of Proposition 2. We just need to show that we can identify the δ_P and γ_P functions since the other functions are identified from Propostion 1. We start with δ_P. The restricted data Q^1 contains $Q_P(p,y|x)$. Since $\Pr(Z > Y|x)$ is observed, the function $R_P(y,p|x) = Q_P(p,y|x)/\Pr(Z > Y|x) = \Pr(Y > y, P > p, Y > P|Z > Y, x)$ is observed. We can then use $R_P(y,p|x)$ and the properties of model (3) to derive the following equality for all $p < y$:

$$\frac{\partial R_P(p,y|x)}{\partial p} = \lambda_P(x)\phi_P(x)\mathcal{L}_G^{(P)}\left(\Lambda_P(p)\phi_P(x), 0, \left[\Lambda_Y(p) + \Omega_P(y,p,x)\right]\phi_Y(x)\right). \tag{B1}$$

The right-hand-side of (B1) is a strictly increasing and identified function of $\Omega_P(y, p, x)$ so $\Omega_P(y, p, x)$ is identified. Then, since λ_P is known, $\delta_P(y, p, x)$ is identified for all $y > p$.

To complete the proof of Proposition 2, we need to show that γ_P is identified. Similarly to what we just did for δ_P, we consider the function $R_{PZ}(p, z) = Q_{PZ}(p, z, -\infty)/\Pr(Y > Z)$ which is available from the restricted data Q^1. We then derive the following equality:

$$\frac{\partial R_{PZ}(p, z|x)}{\partial p} = \lambda_P(x)\phi_P(x)\mathcal{L}_G^{(P)}\left(\Lambda_P(p)\phi_P(x), \left[\Lambda_Z(p) + \Psi_P(z, p, x)\right]\phi_Z(x), 0\right), \qquad \text{(B2)}$$

where the right-hand-side is again a strictly increasing and identified function of $\Psi_P(z, p, x)$ so $\Psi_P(z, p, x)$ is identified and then so is $\gamma_P(z, p, x)$ for all $z > p$. This ends the proof.

Proof of Proposition 3. The proof builds on the same approach as above. Using the functions $Q_Z(z, y|x)$ and $Q_{PZ}(p, z, y|x)$ available from the data Q we can write the two following equalities:

$$\begin{aligned}
\frac{\partial Q_Z(z, y|x)}{\partial z} &= \lambda_Z(z)\phi_Z(x)\mathcal{L}_G^{(Z)}\left(\Lambda_P(z)\phi_P(x), \Lambda_Z(z)\phi_Z(x), \left[\Lambda_Y(z) + \Omega_Z(y, z, x)\right]\phi_Y(x)\right), \qquad \text{(B3)} \\
\frac{\partial^2 Q_{PZ}(p, z, y|x)}{\partial p \partial z} &= \Lambda_P(p)\phi_P(x)\lambda_Z(z)\phi_Z(x)\gamma(z, p, x) \\
&\times \mathcal{L}_G^{(PZ)}\Big(\Lambda_P(z)\phi_P(x), \left[\Lambda_Z(p) + \Psi_P(z, p, x)\right]\phi_Z(x), \\
&\qquad \left[\Lambda_Y(p) + \Omega_P(z, p, x) + \Omega_{PZ}(y, p, z, x)\right]\phi_Y(x)\Big). \qquad \text{(B4)}
\end{aligned}$$

The right-hand-side of (B3) is a strictly increasing and identified function of $\Omega_Z(y, z, x)$ so $\Omega_Z(y, z, x)$ is identified and then so is $\delta_Z(y, z, x)$ for all $y > z$. The right-hand-side of (B4) is a strictly increasing and identified function of $\Omega_{PZ}(y, p, z, x)$ so $\Omega_{PZ}(y, p, z, x)$ is identified and then so is $\delta_{PZ}(y, p, z, x)$ for all $y > z > p$. This ends the proof.