Crépon, B., Ferracci, M., Jolivet, G., & van den Berg, G. J. (2018). Information shocks and the empirical evaluation of training programs during unemployment spells. *Journal of Applied Econometrics*, *33*(4), 594-616. https://doi.org/10.1002/jae.2621

Information Shocks and the Empirical Evaluation of Training Programs During Unemployment Spells

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

July 2017

Abstract

We study the role of notifications in the evaluation of training programs for unemployed workers. Using a unique administrative data set containing the dates when information is exchanged between job seekers and caseworkers, we address three questions. Do information shocks, such as notification of future training, have an effect on unemployment duration? What is the joint effect of notification and training programs on unemployment? Can ignoring information shocks lead to a large bias in the estimation of the effect of training programs? We discuss these issues through the lens of a job search model and then conduct an empirical analysis following a "random effects" approach to deal with selectivity. We find that notification has a strong positive effect on the training probability but a negative one on the probability to leave unemployment. This "attraction" effect highlights the importance of accounting for notifications in the evaluation of active labor market policies.

JEL codes: C31, C41, J64, J68.
Keywords: notification, duration analysis, program evaluation, dynamic treatment assignment, unobserved heterogeneity, policy evaluation.

*CREST-INSEE and J-PAL.
†French Ministry of Labor, Université Paris II and Sciences Po.
‡University of Bristol. E-mail: gregory.jolivet@bristol.ac.uk
§University of Bristol, IFAU Uppsala, IZA, ZEW, J-PAL and CEPR.

A previous version circulated under the title "Dynamic Treatment Evaluation Using Data on Information Shocks". We thank the Editor Ed Vytlacil, three anonymous Referees, Sylvie Blasco, Frank Windmeijer, as well as participants in seminars in Aarhus and Paris and conferences in London, Paris, and EALE/SOLE, for helpful comments.

1 Introduction

The evaluation of active labor market policies (ALMP hereafter) often calls for dynamic approaches. For instance, a relevant policy question is how the exit rate out of unemployment is affected by the date at which the job seeker receives a given treatment, typically a training program. Accordingly, the econometrics literature has taken the standard static evaluation framework with potential outcomes (the Rubin model, 1974) to dynamic settings.[1] A common feature of these approaches is that they rule out unobserved individual information updates about the occurrence of a future treatment. Abbring and Van den Berg (2003a, AVdB hereafter) show that this assumption is crucial for the identification of the treatment effect. Yet in most studies of ALMP's, the individuals' information set prior to training is not available in the data. This raises three important questions for the evaluation of ALMP's. First, do individuals receive information shocks prior to training which may affect their behavior? Secondly, what is the joint effect of training programs and of information shocks on unemployment? Lastly, can the omission of information shocks from the empirical analysis lead to a large bias in the estimation of the effect of training programs? In this paper, we tackle these three questions using a unique administrative data set recording all unemployment and training spells of unemployed workers in Paris in 2003-2004 as well as all the information they receive from caseworkers, in particular notification of future training.

It is useful to interpret the issue we tackle as one of information accumulation over time (see Abbring and Heckman, 2007). If the individual's information set relevant for the future treatment status is fixed over time then inference can proceed in the usual way. However, if individuals receive new information on the future moment of treatment and if they respond to this information, then the econometrician must account for these information shocks. In the case of ALMP's, the existence of such information shocks is plausible. For instance, the caseworker may inform the unemployed worker that he has been assigned to 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). An important feature of our data is that we can observe 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 interpret

[1] A detailed survey of the available techniques is available in Abbring and Heckman (2007). An important reference in this literature is Eberwein, Ham and LaLonde (1997), who were the first to study training effects in a bivariate duration model. Several papers consider approaches based on matching methods without entry selection on unobservables but with conditional independence. An alternative approach is developed by Heckman and Navarro (2007), building on the dynamic discrete-choice literature in combination with instrumental variables.

this notification as an information shock that may affect the job seeker's behavior toward training or unemployment.

To obtain insights into the potential impact of notification on unemployment, we analyze a job search model with endogenous search effort. We show that notification will have an effect on the probability to leave unemployment if training has an effect and if notification affects the chances of starting a training program. The sign of the effect of notification will depend on whether training improves or deteriorates the workers' job prospects. This stylized theoretical model is not taken to the data. Instead, we base our empirical analysis on a reduced-form potential duration outcome model which allows for flexibility regarding heterogeneity and time dependence.

Since we work with observational data, we need to account for selection due to unobserved individual heterogeneity in the reduced-form model. To this end we build on a "random effects" hazard model framework that has been used in a number of empirical studies evaluating ALMP effects on the exit rate out of unemployment. The framework concerns the distribution of the three durations of interest: duration until notification, duration until treatment, and unemployment duration. We allow for individuals to be treated without notification and for treatment dates to be stochastic conditional on notification. We explain in the paper why these two features make sense in the French institutional setting.

An important aspect of our approach is that the analysis of the effect of notification does not require identification of the effect of the treatment on the outcome. We will thus conduct our empirical analysis in two stages. In the first, we consider a partial-information model which leaves aside the evaluation of the treatment effect but focuses on the effects of notification.[2] In the second stage, we estimate a full model and jointly evaluate the effects of notification and treatment.

Whilst we account for observed notifications, we still need to assume that there are no unobserved information shocks at the individual level that affect treatment and/or outcome probabilities. In other words, conditionally on the information received so far, the next shock (information or treatment) cannot be anticipated. One of the innovative contributions of our empirical application will be to check the robustness of our results using data on additional moments where an information flow could have occurred. Our study of anticipation therefore digs at least two levels (notification and additional information shocks) deeper than the literature that ignores anticipation effects. As far as we know, this is the first paper that accounts for so many information layers.

We find that notification has a large positive effect on the probability of being treated and a negative effect on the probability of leaving unemployment. Our results on notification do not hinge on a given specification of the effect of training programs and pass a

[2] As mentioned by AVdB, if the arrival of information is observed, one can redefine the problem as an evaluation of the causal effect of the arrival of information.

series of robustness checks pertaining to the modelling of unobserved heterogeneity, the time-dependence of the notification effect, and the addition of other information shocks. Proceeding to a joint estimation of the effects of notification and training on unemployment, we highlight the importance of accounting for information shocks when conducting an evaluation of training programs. Information shocks such as notification should be seen as part of a richer set of treatments assigned by caseworkers to job seekers. In particular, we show that training policies in France can have a twofold negative effect on exit from unemployment around the date when the training programs start. Indeed, we find evidence of a standard "locking-in" effect at the beginning of the training program and, before the program even starts, of an "attraction" effect whereby the exit rate from unemployment may fall after a notification shock. Training programs do however substantially increase the probability to leave unemployment four months after the start of the program.

A range of existing empirical papers study related topics. Some of these examine the extent to which individuals adjust their behavior in response to knowledge about the moment of future treatments. Black, Smith, Berger and Noel (2003) show that individuals are more likely to leave unemployment once they learn that they must receive compulsory job search assistance. De Giorgi (2005) and Van den Berg, Bozio and Costa Dias (2014) use policy announcement discontinuities to study anticipatory effects of treatments. Lalive, Van Ours and Zweimüller (2005) use Swiss data that contain the moment at which 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. Van den Berg, Bergemann and Caliendo (2009, 2010) show that newly unemployed workers in Germany report widely different subjective probabilities of future participation in ALMP's, including training programs, and that this is reflected in their job search behavior.

The outline of this paper is as follows: section 2 presents the institutional French setting and gives a formal presentation of the role of notifications in the evaluation of training program, first by using a theoretical job search model and then by presenting the statistical model used for estimation. Section 3 describes the content of our administrative data set and discusses the econometric specifications we use for estimation. All the estimation results are in section 4. Section 5 concludes.

2 Training programs with notification

We start this section with a description of the assignment process to training in France and the information that unemployed workers receive when they are notified. The potential effects of these policies are then discussed within the context of a job search model. We end the section with a presentation of the statistical model used in the empirical analysis.

2.1 Training programs and notification procedures in France

Notification. Entry of an unemployed individual 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. 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 contact with a (private or public) training provider.[3] In practice, there are several steps before entering a training program: 1) make a skill assessment with the caseworker; 2) find a training program suited to the needs of the local labor market; 3) find a provider proposing that type of program; 4) find a funding solution for the training.

Passing steps 2 and 3 is not straightforward. This stems from the fact that the full supply of training opportunities is not easily accessible to job seekers, partly because the number of training providers is huge compared to other countries [4]. The lack of a public information system also makes it difficult for a job seeker to find the training program and provider suited to his needs. On the other hand, this information is more accessible to caseworkers, even if there is regional heterogeneity in the quality of the public employment service's information system. Hence, being put in relation with a provider is a crucial step of the assignment process to training. This allows us to define more clearly the nature of the information shock received by the job seekers: some of them will know which provider to contact, while others will not. In theory, the notification is given during, or shortly after, the second meeting with the caseworker. 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 when interviews take place. In the framework of an econometric model, this can be seen as a source of variation in the notification date, which will be supported by descriptive statistics in subsection 3.1.

From notification to training. Once a job seeker is given a notification, he may not immediately enter a training program. In theory, job seekers are free to accept or turn

[3]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.

[4]In France there are more than 60,000 approved training providers, most of them being individual firms, while there are less than 5,000 certified providers in Germany. This is due mostly to the absence of quality assessment in the approval process in France.

down any program they are offered. A refusal can lead to a cut in unemployment benefits, but in practice sanctions for refusing a training program are almost never given.[5] Hence, notification does not imply compulsory training action. This makes the French institutional setting different from systems where sanctions for a refusal of training are much more likely to occur.[6] Moreover, even if the job seeker is eager to be trained, finding a suitable program can take time. This is due to the lack of available training slots or to the time needed to find a funding for the training program. Hence, when notification occurs the job seeker still has to find a funding for her training, which may raise administrative hurdles. Finally, despite recent reforms, the French training system remains complex[7] so notification is only the first step in a possibly long procedure. We will show in subsection 3.1 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, the job seeker has found a training program on his own and this program had to be approved by the caseworker. 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, Jolivet and Van den Berg, 2014). Ideally, we would like to control for the actual content of training programs. Unfortunately, this information is not available in our data so we shall work with a general definition of training programs.[8]

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

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

[7]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, Ferracci and Fougère (2012) for a detailed description of the system.

[8]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. Out of the 593 126 programs that took place between 2005 and 2007, 17.9% were "*general*" (e.g. mathematics, economics, languages), 37.5% were "*personal*" (e.g. development of mental abilities, development of professional organization capacities), 29.9% were "*service oriented vocational skills*" (e.g. accounting, hotel business) and 14.7% were "*production oriented vocational skills*" (e.g. carpentry, engineering).

Additional information shocks. Observing notification of training may not be enough to capture all the information circulating between the caseworker and the unemployed worker. Indeed, prior to the actual notification the caseworker may have sent signals to the worker, during a meeting for instance, that he shall consider some training program. After the notification, the worker may receive further information about the training provider and the content of the program. To address these important concerns, we will consider alternative models to the one where notification is the only information shock.

Our data will allow us to observe the dates of all the ANPE "actions" that is the actions taken by the ANPE caseworker during the job seeker's unemployment spell. These actions could consist of a meeting between a caseworker and a job seeker, in sending a letter to the job seeker, in formally evaluating the worker's skills, in organizing a meeting with potential employers, etc. For instance, authorizing a training program that the job seeker found on his own is an ANPE action and the authorization date will be reported. There are many different types of actions so it will not be possible to model them all separately. What we call notification of training program is a specific ANPE action. We will use these new data to conduct an analysis where we account for three information shocks: notification as well as the first ANPE actions after unemployment starts and (if relevant) after notification.

2.2 Economic interpretation using a job search model

We take a closer look at the main effects at play through the lens of a partial equilibrium model with search frictions, endogenous search efforts, notifications and training programs. This model will not be taken to the data but it will provide some intuition on how notifications may affect the behavior of job seekers and how this effect may mitigate the evaluation of training programs. The following can be seen as an extension of the analysis conducted by Van den Berg et al. (2009), where we introduce notifications and emphasize the role of these shocks in the evaluation of the effect of training.

The environment. Consider a worker who becomes unemployed and receives (constant) unemployment benefits b. Time is continuous and r denotes the interest rate. In this state U, the worker faces three competing risks, all ruled by Poisson processes. He can receive a notification shock, at a rate λ_U^P, go directly to a training program, at a rate λ_U^Z, or receive a job offer, at a rate $\lambda_U^E \cdot s$, where s is the worker's search effort, which we will specify soon. A job offer consists of a job value drawn from a distribution with cdf F^U. For simplicity, we do not model re-entry into unemployment so the value of a job is just the corresponding wage divided by r. We also do not endogenise the decision to participate in a training program so that λ_U^Z is a reduced-form parameter capturing individuals' decision to accept a training program or the extent to which they are forced

to do so (through sanctions on benefits for instance).

If an unemployed worker receives a notification shock, he is in state P and faces two competing risks, also ruled by Poisson processes. He can start a training program at rate λ_P^Z or receive a job offer, at a rate $\lambda_U^E \cdot s$ where s denotes the job search effort. We assume that notified workers also draw their job offers from the distribution F^U.

If a worker starts a training program, state Z, the only shocks he faces are job offers, arriving at a Poisson rate $\lambda_Z^E \cdot s$, and drawn from a distribution with cdf F^Z. The following analysis would still hold if we allowed for a locking-in effect i.e. a fixed time period during which workers who start a training program cannot receive job offers.

Workers reject job offers with a value below their current value. Workers choose the search efforts which maximize their value function in each state (U, P or Z). We have already specified the returns-to-search technology ($\lambda \cdot s$) and we assume that a search effort of s generates a flow cost of $c \cdot s^2/2$, where $c > 0$ is a constant parameter.

In what follows, we will often refer to a pair (λ, F) as a search environment (arrival rate of job offers and distribution in which they are drawn).

Reservation values and search efforts. Consider a trained job seeker and let $V_Z(s)$ be his expected utility if he searches with effort s. Under our assumptions, we can define its maximum, denoted as V_Z, and the corresponding search effort s_Z. We can derive the value and search effort for states U (V_U and s_U) and P (V_P and s_P) in a similar fashion. We skip the details of the calculations[9] and write down the dynamics of the three value functions:

$$rV_Z = b + \frac{1}{2c}\left[G_Z\left(V_Z\right)\right]^2, \tag{1}$$

$$rV_P = b + \frac{1}{2c}\left[G_U\left(V_P\right)\right]^2 + \lambda_P^Z\left(V_Z - V_P\right), \tag{2}$$

$$rV_U = b + \frac{1}{2c}\left[G_U\left(V_U\right)\right]^2 + \lambda_U^Z\left(V_Z - V_U\right) + \lambda_U^P\left(V_P - V_U\right), \tag{3}$$

where $G_U(V) = \lambda_U^E \int_{v \geq V}(v - V)dF^U(v)$ and $G_Z(V) = \lambda_Z^E \int_{v \geq V}(v - V)dF^Z(v)$ can be seen as the expected gains of a worker with value V searching with effort $s = 1$ in the search environment of non trained workers (G_U) or of trained workers (G_Z). Note that these two functions are decreasing.The optimal search efforts are characterized by:

$$s_U = G_U\left(V_U\right)/c, \qquad s_P = G_U\left(V_P\right)/c, \qquad s_Z = G_Z\left(V_Z\right)/c. \tag{4}$$

We note that the search effort in a given state decreases as the value of this state increases.

[9]They are relatively straightforward. We start from the Bellman equation of the value of a trained worker with search effort s: $rV_Z(s) = b - cs^2/2 + \lambda_Z^E \cdot s \cdot \int_{V_Z(s)}\left[v - V_Z(s)\right]dF^Z(v)$, then maximize $V_Z(s)$ with respect to s to get s_Z and V_Z. We then proceed similarly with the value of a notified worker with effort s and so on.

In this paper, we are mainly interested in transition probabilities i.e. in hazard rates. In this search model, the two main determinants of a worker's transition to employment are the value he attaches to his current state and his search effort. The latter will drive the arrival of job offers and the former will lead him to accept or reject an offer. The values V_U, V_P and V_Z are thus the job seeker's reservation values in each state. Having characterized the value functions and search efforts, we can now write the hazard rates out of each of the three states U, P and Z, into employment. They are given by:

$$h_U^E = \lambda_U^E s_U \left[1 - F^U\left(V_U\right)\right], \qquad h_P^E = \lambda_U^E s_P \left[1 - F^U\left(V_P\right)\right], \qquad h_Z^E = \lambda_Z^E s_Z \left[1 - F^Z\left(V_Z\right)\right]. \tag{5}$$

We note that these hazard rates unambiguously decrease with the value of the state of origin. For instance, h_U^E decreases with V_U as it is the product of two positive and decreasing functions of V_U: s_U and $1 - F^U$. Also, in this job search model, we abstract from time-dependent hazard rates to keep the model stationary. The empirical analysis will allow for time dependence.

Discussion: effect of notification. We now use this job search model to delve into the main issues arising from the presence of notifications. We start with the effect of training. In this model, training can change a job seeker's employment prospects by increasing the arrival rate of job offers, if $\lambda_Z^E > \lambda_U^E$, or by improving the quality of the job offers. This would be the case if the worker's productivity increased during the training program, allowing him to apply to better paid jobs and, formally, it could be reflected by stochastic dominance of F^Z over F^U. In this case, we will say that training improved the worker's search environment and our model captures this formally by having $G_Z > G_U$ (returns to search are higher for trained workers). On the contrary, it could be that training deteriorates a worker's search environment, for instance if the locking-in effect is so large that trained workers receive fewer offers and the pool of jobs they apply to is not better (formally, if $\lambda_Z^E < \lambda_U^E$ and $F^Z = F^U$). Our model captures this deteriorating effect when $G_Z < G_U$. Lastly, it could be that training has no effect: $G_Z = G_U$.

If training does improve the worker's search environment, one can show that $V_Z > V_P$, using (1), (2) and (5). Looking at (2) and (3), this will lead to a difference between the values of states U and P if $\lambda_P^Z > \lambda_U^Z$. In this case, notification increases a worker's chances of going to a stage, Z, where his job prospects are better. Then $V_P > V_U$ and, using (5), $h_P^E < h_U^E$. If however, $\lambda_P^Z = \lambda_U^Z$ then the positive effect of training does not generate any difference between the states U and P, and thus notification does not affect a worker's hazard rate.

This discussion illustrates a key feature for our analysis. We are mainly interested in the effect of notification on unemployment duration. Our job search model tells us that there will be such an effect if two conditions are verified: training must have an effect on workers' job prospects, formally $G_Z \neq G_U$, and notification must change a job seeker's

probability to be treated, formally $\lambda_P^Z \neq \lambda_U^Z$. Should one of these channels be missing, notification will have no impact on unemployed workers' hazard rate.

Taking stock of the effect of notification, we have the following formal results:

- If $G_Z > G_U$ and $\lambda_P^Z > \lambda_U^Z$ then $V_P > V_U$ and $h_P^E < h_U^E$.
- If $G_Z < G_U$ and $\lambda_P^Z > \lambda_U^Z$ then $V_P < V_U$ and $h_P^E > h_U^E$.
- If $G_Z = G_U$ or $\lambda_P^Z = \lambda_U^Z$ then $V_P = V_U$ and $h_P^E = h_U^E$.

Discussion: treatment evaluation in the presence of notification. A relevant question for the evaluation of training programs is how they affect the hazard into employment. This will be captured by the difference:

$$\Delta_{ZU} = h_Z^E - h_U^E. \tag{6}$$

As mentioned earlier, the hazard rates, and thus the treatment effect, are constant in this job search model. This will not be the case in the empirical analysis. If we did not observe notification and compared the hazard rates of trained and non-trained workers, we would measure:

$$\Delta(t) = \Delta_{ZU} + \omega(t) \cdot \Delta_{PU}, \tag{7}$$

where $\Delta_{PU} = h_P^E - h_U^E$ is the change in the hazard rate into employment due to notification and where $\omega(t)$ if the probability of being notified before date t conditionally on being still unemployed and not having started training before t. Using the Bayes' rule and the Poisson structure of the shock processes, we can write $\omega(t)$ as a function of t and of the hazard rates:[10]

$$\omega(t) = 1 - \frac{1}{1 + \lambda_U^P \cdot \left[1 - e^{\left(\lambda_U^P + \lambda_U^Z + h_U^E - \lambda_P^Z - h_P^E\right)t}\right]}. \tag{8}$$

Equation (7) shows that the difference between the hazard rates of trained and non-trained workers ($\Delta(t)$) differs from the effect of training on unemployed workers (Δ_{ZU}) if notification affects the probability to leave unemployment (if $\Delta_{PU} \neq 0$). This difference becomes larger when the proportion $\omega(t)$ of notified workers among unemployed, non-trained individuals increases. As shown above, notification has no effect on unemployment duration if training has no effect on exit or if the probability of being treated does not change with notification. In this case $\Delta_{PU} = 0$ and studies using the standard evaluation framework with no notifications will estimate Δ_{ZU}. If notification has an effect on unemployment duration, the group of workers not-yet trained at a given date is heterogeneous with respect to their hazard rate into employment, which will create a difference between these studies' estimation target and the effect of training (Δ_{ZU}).

[10] The hazard rates are either exogenous parameters (for transitions to notification or training) or, for transitions to employment, functions of the endogenous search effort and value function.

Equation (8) illustrates how some parameters may affect the composition of the group of non-trained workers. For instance, if the hazard of starting training when notified, λ_P^Z, increases, then there is a direct negative effect and an indirect effect on $\omega(t)$. The direct effect comes from the fact that notified workers go more quickly into training so their proportion in the group of non-trained workers decreases. The indirect effect is coming from h_P^E. Indeed, notified workers' access to training changes so, if training has an effect ($G_Z \neq 0$), the value V_P will change and so will the probability to exit to employment. If $G_Z > 0$ (if training helps workers find better jobs more quickly), this indirect effect is positive so the overall effect of λ_P^Z on $\omega(t)$ is ambiguous.

We thus need to conduct an empirical analysis to quantify the effects of notification on unemployment duration and on the evaluation of training. In the next subsection we present a reduced-form model allowing for individual observed and unobserved heterogeneity as well as time dependence of the hazard rates. Hence, the empirical analysis does not involve a structural estimation of the above job search model (see the discussion in Van den Berg, 2001).

2.3 The statistical model

We briefly present the statistical model we use for the evaluation of training programs in a dynamic setting with notifications. In particular, we discuss the three main issues of interest, pertaining to the effect of notifications on unemployment duration, the effect of training on unemployment duration and how the evaluation of this latter effect may be affected by the workers' response to notifications.

Potential durations. We want to evaluate the effect of a treatment (training programs) on the duration an individual spends in a state of interest (unemployment). The treatment can be assigned at different points z in time. We let Z denote the duration until treatment and Y the duration in the state of interest (also called the outcome). The individual can receive information about future training. More specifically, he can receive a notification, from the caseworker, that he is likely to start a training program in a near future.[11] We define hypothetically assigned moments p of information arrival, and a corresponding random variable P, denoting the duration from $t = 0$ until the actual moment of arrival of this information. We do not specify how precise the information is concerning the future moment of training. For now we consider at most one information package per individual. In an extension later in this paper we allow for subsequent information packages.

In the model, the notification arrives at a certain rate. We follow this route as it allows us to use a simple extension of the standard ToE setting and also because it

[11] We give more details on the nature of this information shock in section 3, where we present the actual institutional setting.

is difficult to observe in the data details of the contents of the notification. Should this information be available, a more realistic representation of notifications would model them as heterogenous shocks to the worker's information set. Note that essentially the same issue arises in the evaluation literature on training programs, as the details of the contents of training are often only partially observed by the econometrician.

We can now define the potential-outcome durations of interest. First, $Z(p)$ is the time until training if he is assigned to a notification at time $t = p$ (both measured from the moment of entry into unemployment). Further, $Y(z, p)$ is the unemployment duration if the individual is assigned a treatment at date $t = z$ and is assigned a notification at date $t = p$. The durations $Z(p)$ and $Y(z, p)$, where $p, z \geq 0$, are potential-outcome random duration variables.

We consider vectors of observed covariates X and unobserved (to the econometrician) covariates V that, together, are systematic determinants of outcomes and/or treatment assignment and/or notification. We can then define the potential outcomes and the actual treatment for any given (X, V), and we can subsequently model selection effects as effects of (X, V).

Three main issues of interest. We can now present the three main questions that we aim to address in this paper. The first one pertains to the effect of notifications on unemployment duration. More precisely, we want to know if the process driving unemployment duration changes once the individual has been notified i.e. do unemployed workers change their job search strategy once they have been told that they may soon start a training program?

It is important to explain how this issue relates to anticipation of future treatments. Workers may know that they will start a training program with some positive probability and set their search effort and strategy accordingly. Such "ex ante" behavior would not violate the identifying assumptions of the standard dynamic treatment literature that followed AVdB. In this paper, we want to know whether job seekers respond to the arrival in time of new information about future treatments, by changing their job search and acceptance behavior. The job search model derived in the previous section sheds light on this important distinction.

The second main issue of interest is more policy-oriented and revolves around the effect of training programs on unemployment duration. This is usually the main estimation target in empirical studies that use dynamic evaluation frameworks to study ALMP's. Importantly, our analysis of the effects of training will be conducted in a setting where job seekers can receive prior notification of the treatment, which may affect unemployment duration even before training actually starts. This raises the question of which counterfactual to use when assessing the effect of the treatment. We will consider both cases i.e. the effect of training on the unemployment duration of a notified job seeker and of a

non-notified one.

This takes us to the last issue of interest: can we assess the effect of training programs if we ignore, or do not observe, notifications? In most studies, the exchange of information between caseworkers and job seekers is not observed, which forces econometricians to assume that no unobserved shocks affect unemployment duration prior to the treatment. If notification triggers a change in the behavior of job seekers, this assumption is violated and this will affect the evaluation of the treatment. More precisely, if notification is unobserved, the evaluation of the treatment will rest on the comparison of the hazard rate of a treated individual and that of a not-yet-treated individual, who may or may not be notified.

Reduced form specification. We impose mixed proportional (MPH) hazard rates on the duration processes of interest. Let X be a vector of observed individual characteristics and $V = (V_P, V_Z, V_Y)$ be a vector of unobserved individual characteristics, independent of X. The hazard rates at date t and conditional on (X, V) are denoted as $h_P(t|X, V)$ for P, $h_Z(t|p, X, V)$ for $Z(p)$ and $h_Y(t|z, p, X, V)$ for $Y(z, p)$. We specify:

$$\begin{aligned} h_P(t|X,V) &= \lambda_P(t)\phi_P(X)V_P, \\ h_Z(t|p,X,V) &= \lambda_Z(t)\phi_Z(X)V_Z \cdot \exp\left[\gamma_P(t,p,X)\cdot \mathbf{1}\left\{p<t\right\}\right], \\ h_Y(t|z,p,X,V) &= \lambda_Y(t)\phi_Y(X)V_Y \cdot \exp\left[\delta_P(t,p,X)\cdot \mathbf{1}\left\{p<t\leq z\right\} + \delta_Z(t,z,X)\cdot \mathbf{1}\left\{z<t\right\}\right]. \end{aligned} \tag{9}$$

The functions δ_P, δ_Z and γ capture the effect of notification on the reemployment rate, the effect of training on the reemployment rate, and the effect of notification on the rate of entering training. The λ and ϕ functions are specified below; the λ functions should not be confused with the structural λ parameters in subsection 2.2. Imposing $\gamma_P = \delta_P = 0$ and ignoring h_P yields the standard ToE model of AvdB which has been used in many evaluation studies.[12]

Two key implications of the modeling of notifications and treatments are that: *i*) one can enter treatment without having received an information shock (Z can be smaller than P) and *ii*) the starting date of the treatment is still random after P has been realized (the distribution of Z given P is not degenerate). These characteristics of the model are introduced with an eye on our empirical application. One could also specify a slightly modified model in which the information shock necessarily arrives before the treatment.

Notice that model (9) rules out effects of anticipation of the moment of notification. Indeed, notification can have an effect on the duration until training (Z) or on the unemployment duration (Y) only after the job seeker has been notified by the caseworker. Likewise, training can affect the exit rate to work only after the start of the training program. In model (9), these features are captured by the indicator functions in (9).[13]

[12] We need to make 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).

[13] In a more general setting, with no functional form assumptions on the hazard rates, we would have

The model also makes a conditional independence (CIA) assumption. Specifically, conditionally on X and V, $Y(z,p)$ is independent of $(Z(P),P)$ and $Z(p)$ is independent of P. Notice that this CIA is unconventional in the sense that it conditions both on observed and on unobserved covariates. This is because we allow for selection due to unobserved random effects capturing unobserved confounders. The proportionality assumptions in the model are also important for identification and are discussed in subsection 3.2 below. If one can rule out the presence of unobserved confounders, then the usual CIA assumption applies, and, in addition, the proportionality assumptions can be relaxed and one could adopt a dynamic matching approach (see e.g. Crépon, Ferracci, Jolivet and Van den Berg, 2009).

The set of functions $(\gamma_P, \delta_P, \delta_Z)$ describes the effects of information shocks and treatment on the durations of interest. The three issues of interest discussed above revolve around the value of δ_P (capturing the effect of notification on unemployment duration), δ_Z (capturing the effect of training on unemployment duration) and how the evaluation of the training effect δ_Z is affected by taking notifications P into account. The theoretical analysis in subsection 2.2 suggests that the first and third issues also depend on the effect γ_P of notification on the rate of being trained.

3 Empirical application

3.1 Data and descriptive statistics

The data set. Our data come from the Fichier Historique Statistique (FHS hereafter), an exhaustive register of all unemployed spells recorded at the ANPE, whether the individual receives unemployment benefits or not. We use data on all unemployment spells in the city of Paris and starting in 2003 or 2004. We follow these spells 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[14] 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 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

to rule out unobserved individual-specific shocks prior to P that can affect the treatment probability, unobserved individual-specific shocks prior to P and Z that can affect the outcome and unobserved individual-specific shocks prior to treatment that (conditionally on the realization date of the information shock) can affect the outcome. A formal statement of these assumptions in the context of a potential duration model is available upon request.

[14] 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.

towards a training program. As explained in subsection 2.1, we consider that a job seeker has received notification of a future treatment when he is informed by the caseworker that he shall be put in contact with a training provider. Lastly, as we discussed at the end of the same subsection, we observe the dates of all ANPE "actions" that is the day when a meeting takes place between the caseworker and the job seeker or when a letter is sent to a job seeker. This information will be useful to conduct robustness checks.

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 duration until notification P_i if $C_i^P = 0$ but we only know that this duration is longer than P_i if $C_i^P = 1$. We observe the duration until treatment Z_i if $C_i^Z = 0$ but we only know that this duration is longer than Z_i if $C_i^Z = 1$. We observe the unemployment duration Y_i if $C_i^Y = 0$ but 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. These are denoted by the vector X_i and consist of: $\mathbf{1}\{\text{male}\}$, 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") and education (6 categories, the reference is "university degree"). Lastly, we use the location of the unemployment agency to define an individual's local labor market and we compute two indicators of the latter. 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 483,523 unemployment spells, starting between the 1st of January 2003 and the 31st of December 2004. Only 4.50% of these spells are censored by the data extraction date (1st of January 2008). 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 or age. We note that relatively few individuals are notified (9%) or trained (8%), 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. Note that our modeling of the hazard rates for P and Z, cf. model (9), is consistent with the statistics shown in Table 1, in particular with $\Pr(Z < P) > 0$.

Table 1: Fractions of spells receiving notification and/or training

	all sample	male	female	age ≤ 25	age ≥ 55
% notified	9	7.9	10.2	6.9	3.5
% treated	7.9	7	8.9	5.5	2.5
% treated if not notified	5.7	5	6.5	4.2	1.7
% treated if notified	30.2	30.4	30.1	23.1	25.3

Table 2 shows the average and a series of quantiles for observed durations of interest. Unemployment spells have an average duration of almost 11 months. Note that there is a lot of variation in the date when notification is given, with an average of about 6 months (which is consistent with the interview process introduced by the PARE reform). There is also variation in the starting date of training programs, with an average of about 8 months (233 days). For those who were given notification and actually started a training program, the interval between these two events is around 3 months on average.

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

	Mean	Q10	Q25	Q50	Q75	Q90
P if notified	172	8	24	93	238	447
Z if treated	233	40	92	184	329	502
Z if treated and not notified	215	31	81	166	302	474
Z if treated and notified	269	64	118	220	378	544
$Z - P$ if treated and notified	95	6	20	55	119	242
Y	328	28	64	202	475	834
Y if not notified and not treated	288	26	54	165	386	780
Y if notified and not treated	489	76	190	384	740	1,030
Y if notified and treated	669	282	422	657	854	1,071

3.2 Inference

For each individual in the data, we observe X and Y, although the latter can be censored by the sampling date.[15] 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 leaves before having been treated, we only know that $Z \geq Y$. Likewise, we observe P if and only if

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

$P < \min(Z, Y)$. If an individual starts treatment or leaves the state of interest without having received the information shock, we only know that $P \geq \min(Z, Y)$.

Using these data, we estimate the reduced-form models described in subsection 2.3. Specifically, we estimate them as random effects models, i.e. by invoking Maximum Likelihood Estimation, integrating over the distribution of the unobserved heterogeneity terms. The discussion of this can be brief. The identification of the models follows straightforwardly from the identification proofs in the literature (see e.g. AVdB, Abbring, 2008, and Abbring and Heckman, 2007). In subsection 2.3 we already mentioned two important conditions for identification, namely a conditional independence assumption (relaxing the usual CIA assumption by conditioning on unobservables as well as observed covariates) and an assumption ruling out anticipation (in particular, in our setting, ruling out anticipation of the notification date). Two other assumptions are important. First, we require the "random effects" assumption that, in the inflow into unemployment, unobserved covariates are independent of observed covariates. We hope to accommodate for this to some extent by including as many observed covariates as possible, to the boundary of what is computationally feasible. Secondly, we require the hazard rates to follow MPH-type specifications. Like in most cases where additivity and proportionality assumptions are made, it is difficult to justify this assumption economically. As shown by Abbring and Van den Berg (2003b), the most important aspect of the MPH assumption in this context is that the hazard rates are proportional in the unobserved covariates. Intuitively, the latter ensures that the selective treatment assignment creates a global statistical dependence that is present at all durations. Conversely, the causal treatment effect creates a local dependence as it only works from the moment of treatment onwards. If the realization of the duration outcome of interest is typically shortly preceded the treatment, then this is evidence of a causal effect of training. The spurious selection effect does not give rise to the same type of quick succession of events. From this it is obvious that the use of the proportionality assumptions is particularly problematic if in reality there are unobserved shocks that affect both the treatment rate and the rate at which the outcome of interest occurs. Notifications and meetings with case workers are examples of shocks that affect the training rate and the exit rate to work. Hence, in evaluation settings with MPH specifications for the hazard rates, it is important that such shocks are observed and that the model includes them. This is of course exactly what we do in our analysis of information shocks, and this provides an additional motivation for this analysis.

It is important that functions that act as model determinants have flexible forms. We now discuss these functional form specifications. First we consider a "partial" model which exploits data only until individuals start a training program or leave unemployment untrained. Again, as follows directly from the literature, the effects of notification are identified from these data, i.e., are identified without making assumptions about the δ_Z function i.e. about the effect of training on unemployment.

The duration model. We use the K_P-quantiles of P conditionally on $C^P = 0$ as cut-off points for the piecewise constant part of the hazard rate in (9). This introduces $K_P - 1$ parameters to estimate for λ_P as, for normalization, we fix the probability on the first interval, λ_{P1}, to be .0001.[16] We proceed similarly for λ_Z and λ_Y (except that we do not condition on $C^Y = 0$ for the latter). We set $K_P = K_Z = K_Y = 11$. The 30 parameters thus introduced are stacked in the vector Λ. The ϕ functions in (9) are specified as log-linear functions: $\phi_P(X) = \exp(X'\beta_P)$, $\phi_Z(X) = \exp(X'\beta_Z)$ and $\phi_Y(X) = \exp(X'\beta_Y)$.

Effects of notification. We estimate two specifications for the effects of notification on training, γ_P, and unemployment, δ_P. The first specification simply models these effects as constant. The second specification allows these effects to change after one or after three months. For instance γ_P satisfies: $\gamma_P(t, P, X) = \gamma_P^0 \cdot \mathbf{1}\{t \leq P + 30\} + \gamma_P^1 \cdot \mathbf{1}\{P + 30 < t \leq P + 90\} + \gamma_P^3 \cdot \mathbf{1}\{P + 90 < t\}$. In this latter specification, γ_P and δ_P thus refer to vectors of three parameters. In the "partial" model do not specify the treatment effect δ_Z.

The distribution of unobserved heterogeneity. The distribution of unobserved heterogeneity is assumed to have a discrete support with a given number R of mass points. More precisely $\Pr(V = \exp(v_r)) = p_r$, $\forall r \in [1, R]$, where $v_r \in (0, \infty)^3$. The probabilities are modeled as: $p_r = \frac{\exp(-\alpha_r)}{\sum_{r=1}^{R} \exp(-\alpha_r)}$, where $\alpha_R = 0$ and $\alpha_r \in \mathbb{R}$ if $r < R$.

For a given number R, 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 pairs (if there are two processes, triplets if there are three, etc.) of these values. Our approach, which follows McCall (1996), is convenient if there are more than two duration processes (as the number of parameters to estimate increases more slowly when we account for more groups).

The likelihood. We stack γ_P, δ_P, Λ, β_P, β_Z, β_Y and $\{p_r, v_r\}_{r\in[1,R]}$ in the vector Θ_R. As mentioned above, when estimating the "partial" model, we follow individuals only up to the minimum of Z and Y and record which of these two durations is the shortest. Hence C_i^Y is now equal to 1 if $Z_i < Y_i$. The contribution to the likelihood of spell i is given by:

$$\begin{aligned}\ell\ \left(P_i, C_i^P, \min(Z_i, Y_i), C_i^Z, C_i^Y | X_i, \Theta_R, R\right) = \sum_{r=1}^{R} p_r \left[h_P(P_i|X_i, v_r)\right]^{1-C_i^P} S_P(P_i|X_i, v_r) \\ \times \left[h_Z(\min(Z_i, Y_i)|X_i, v_r)\right]^{1-C_i^Z} S_Z(\min(Z_i, Y_i)|X_i, v_r) \\ \times \left[h_Y(\min(Z_i, Y_i)|X_i, v_r)\right]^{1-C_i^Y} S_Y(\min(Z_i, Y_i)|X_i, v_r), \qquad (10)\end{aligned}$$

[16] In ToE models, the MPH structure implies that the λ's are identified up to scale.

where the hazard rates h_P, h_Z and h_Y are given by (9) and S_P, S_Z and S_Y denote the corresponding survival functions. For instance $S_P(t|X,V) = \exp\left[-\int_0^t h_P(u|X,V)du\right]$.[17] Note that the individual values of the v_r are random effects over which we integrate the individual likelihood contributions.

We write the full likelihood function conditionally on the number of groups of unobserved heterogeneity as follows:

$$L\left(\Theta_R, R\right) = \prod_{i=1}^{N} \ell\left(P_i, C_i^P, Z_i, C_i^Z, Y_i, C_i^Y | X_i, \Theta_R, R\right). \tag{11}$$

Alternative specifications. The simplest specification above has constant effects γ_P and δ_P, and $R = 2$ groups of unobserved heterogeneity. We call this the baseline model specification. To check for robustness, we will consider three alternative specifications. First, we allow the effects of notification, γ_P and δ_P, to vary over time. Then we estimate a model with more groups of unobserved heterogeneity: $R = 3$ or $R = 4$. Next, we assess the assumption that workers do not act on the arrival of information that is not captured by the notification P. To this end, a third alternative specification modifies the baseline specification by exogenously censoring $\min(Z, Y)$ after 30 days once notification has been received. This serves to make the estimation of γ_P and δ_P more robust to the arrival of information shocks posterior to notification – as long as these shocks are not realized within a month. Our data allow us to see how long it takes for an ANPE action to be realized once the worker has been notified of a future treatment. We find that only a small proportion of actions, 12%, are realized within 30 days so not using information older than a month after notification should yield estimates that are immune to anticipation based on posterior information shocks.

If there are other shocks than P that may affect the job seeker's hazard rates, we can also use the additional information contained in our data on these shocks. Let A be the duration from the start of the unemployment spell until the first ANPE action prior to notification and let B be the duration between notification and the first ANPE action posterior to notification. We make MPH assumptions on the hazard rates of A and B and include them in model (9) as two additional duration processes. The hazard rate of P may then include a multiplicative effect η_A capturing the change in the arrival rate of

[17] In fact, our data contain time-varying elements of X_i, capturing the local labor market conditions of the individual. For ease of exposition we have ignored time-varying covariates in the exposition on expressions for the empirical models and the likelihood function. As argued by Eberwein, Ham and LaLande (1997), time-varying covariates provide particularly useful information on parameters of models with different duration variables. Heuristically, a comparison of individuals whose covariates change to those whose covariates are constant helps to detect causal effects of the covariates (see formal identification proofs in e.g. Honoré, 1991, and Brinch, 2007). In addition, if an exclusion restriction can be made then the association of time variation in a covariate of one hazard rate, with a duration variable governed by another hazard rate, indicates the presence of correlated unobserved heterogeneity.

notification after the first ANPE action A. Likewise, as long as P is not yet realized, A can have an effect on Z, denoted as γ_A, or, as long as P or Z is not yet realized, on Y, denoted as δ_A. The post-notification shock B can also have an effect on Z, denoted as γ_B, or, as long as Z is not yet realized, on Y, denoted as δ_B.[18]

It is interesting to see how the estimates of δ_P and γ_P change with respect to the baseline model. We can also look at the estimates of the effects of A and B on the training probability and on unemployment duration and then assess whether notification is the only relevant shock. Ideally, we should model all the ANPE actions, i.e. all the information shocks, not only A, P and B. Whilst this is theoretically possible, in practice estimating a duration model with so many processes raises numerical problems that are beyond the scope of this analysis.

Effect of training. The only feature of interest that does not appear in the "partial" model is the effect of training programs on unemployment, δ_Z. In the analysis of the full model, we follow individuals beyond entry into training. Formally, the duration Y is no longer censored by Z. The contribution to the likelihood of an individual i is then given by:

$$\begin{aligned}\ell\left(P_i, C_i^P, Z_i, C_i^Z, Y_i, C_i^Y | X_i, \Theta_R, R\right) = \sum_{r=1}^{R} p_r & \left[h_P(P_i|X_i, v_r)\right]^{1-C_i^P} S_P\left(P_i|X_i, v_r\right) \\ \times & \left[h_Z\left(Z_i|X_i, v_r\right)\right]^{1-C_i^Z} S_Z\left(Z_i|X_i, v_r\right) \\ \times & \left[h_Y\left(Y_i|X_i, v_r\right)\right]^{1-C_i^Y} S_Y\left(Y_i|X_i, v_r\right), \quad (12)\end{aligned}$$

where the hazard rates are as in (9). We allow the treatment effect δ_Z to change after 4 or 12 months. Formally: $\delta_Z(t, Z, X) = \delta_Z^0 \cdot \mathbf{1}\{t \leq Z + 120\} + \delta_Z^4 \cdot \mathbf{1}\{Z + 120 < t \leq P + 365\} + \delta_Z^{12} \cdot \mathbf{1}\{Z + 365 < t\}$.

4 Results

4.1 The effect of notification on training and unemployment

We first focus on the estimates of the effect of notification on the duration until training and on unemployment duration. The dependence of hazard rates on individual heterogeneity and on time will be discussed in subsection 4.2. All estimates presented in the

[18] The hazard rates of Z and Y are modeled as follows:

$$\begin{aligned}h_Z(t|B, P, A, X, V) &= \lambda_Z(t)\phi_Z(X)V_Z \cdot \exp\left[\gamma_A \mathbf{1}\{A \leq t < P\} + \gamma_P \mathbf{1}\{P \leq t\} + \gamma_B \mathbf{1}\{P + B \leq t\}\right], \\ h_Y(t|Z, B, P, A, X, V) &= \lambda_Y(t)\phi_Y(X)V_Y \cdot \exp\left[\delta_A \mathbf{1}\{A \leq t < \min(P, Z)\} + \delta_P \mathbf{1}\{P \leq t < Z\}\right. \\ &\quad \left. + \delta_B \mathbf{1}\{P + B \leq t < Z\} + \delta_Z \mathbf{1}\{Z \leq t\}\right].\end{aligned}$$

current subsection 4.1 are based on the partial-information model, where we do not follow individuals after they start a training program or leave unemployment. Hence these estimates do not depend on a specification of the treatment effect of training programs. The joint evaluation of the effects of notification and training will be discussed in subsection 4.4.

We start with the baseline specification (see subsection 3.2). The estimates are in Table 3.[19] We see that notification has a significant, large and positive effect on the treatment probability. Once P is realized, the hazard rate of Z increases by a factor $\exp(3.555) \approx 35$. This is expected since notification P is the first step towards a state-sponsored training program. Hence, the information shock we consider does convey information on the treatment probability.

Table 3: Effects of notification — baseline specification

Effect on training (γ_P)	3.555	$(0.037)^{***}$
Effect on exit from unemployment (δ_P)	-0.370	$(0.011)^{***}$

There is no one-to-one relationship between notification and treatment status. 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 3 show that the probability to enter a training program is much higher once an individual has received a notification.

The estimate of δ_P shows that notification also significantly lowers the probability to leave unemployment. The hazard rate of Y decreases by 30% ($\approx 1 - \exp(-0.37)$). This goes against a "threat effect" of notification (see, e.g., Black et al., 2003). The latter would work as follows. If workers who are notified dislike participating in a training program (e.g. because it takes time to participate), they might leave unemployment for a job that they would not have accepted otherwise. We do not find such an effect in our data. Instead, we find an "attraction" effect whereby workers are less likely to leave unemployment once they receive notification of a future treatment.

We can interpret this attraction effect using the job search model derived in subsection 2.2. In this model, we showed that the probability to leave unemployment decreases

[19] In all tables shown in this section, standard errors are in parenthesis and one, two and three stars denote significance at the 90%, 95% and 99% level respectively.

after notification if notification increases the probability to start a training program and if training improves workers' search environment (i.e. trained workers obtain more and/or better job offers). Along these lines, and following Van den Berg et al. (2009), one could even infer the sign of the training effect from the estimates in Table 3. Given that notification increases the rate to be treated, notified workers would leave unemployment faster to avoid training in case the latter is ineffective or unattractive. Since, in contrast, we observe a reduction of the rate of leaving unemployment after notification, it follows that training is valued positively by job seekers. Assuming that the time in training is not valued positively, it follows that training has a positive ex post effect.[20]

In short, and this is one of the main results of our paper, workers do act upon the realization of private signals informative about their future treatment status. Whilst we will take a closer look at the evaluation of training programs in subsection 4.4, it already follows that, in our application, the assumptions of a standard ToE evaluation model that does not account of notifications are rejected by the data.

Robustness checks We now check the robustness of our results on notification to changes in the model specification, apart from the extension accounting for additional observed information shocks before and after notification. The latter will be discussed in subsection 4.3.

The first panel of Table 4 reports estimates from a specification where we allow for the effects of notification to change after one month and after three months. As one may expect, the effect of notification on treatment is stronger in the first weeks and then falls after three months but remains large and significant. For the effect of notification on the outcome Y we find a steep increasing pattern through time. Compared to its pre-notification level, the hazard rate of Y is 45% (≈1-exp(-0.603)) lower in the first month following notification, 26% lower in the next two months and 19% lower later on. Hence, the effects of notification on treatment and outcome vary over time and tend to be stronger, in absolute value, right after notification. They are still significant more than three months after notification.

The second panel of Table 4 shows estimates for the baseline specification where Z and Y are censored after 30 days after notification. These estimates should be robust to the presence of additional information shocks arriving after notification as long as the realization of these shocks takes more than one month. In our data, we observe that only 12% of the ANPE actions realized after notification have taken place in less than 30 days after notification. We find a slightly higher point estimate for γ_P and a much

[20]The possibility to receive a sanction for refusing to start a training program may lead to a positive (threat) effect of notification on the hazard rate of Y, provided that sanctions are imposed in practice with a sufficiently large probability. As we have seen, in France, sanctions for refusing training programs are rarely implemented.

lower estimate for δ_P. Once again, there is strong evidence of that workers respond to notifications.

Table 4: Robustness checks: alternative specifications

	γ_P		δ_P	
Time-varying effects:				
- first 30 days	3.487	(0.065)***	-0.603	(0.029)***
- next 60 days	3.204	(0.085)***	-0.454	(0.024)***
- after 90 days	2.471	(0.102)***	-0.206	(0.019)***
Censoring durations 30 days after notification:				
	3.937	(0.060)***	-0.663	(0.020)***
Changing the distribution of unobserved heterogeneity:				
$R = 2$	3.555	(0.037)***	-0.370	(0.011)***
$R = 3$	4.717	(0.075)***	-0.396	(0.013)***
$R = 4$	3.691	(0.082)***	-0.347	(0.011)***

The difference in this estimate of δ_P and the one in Table 3 can easily be explained by referring to the first panel of Table 4, which allowed for the notification effects to vary over time. We actually note that the δ_P-estimate found in the second panel of 4 is close to the one we found in the first panel for the first month after notification. This makes sense since the estimate in the second panel δ_P is constant over time but does not use the information arriving later than a month after notification. Hence, the first conclusions we draw from the first or second panels of Table 4 are similar, in that the effects of notification seem to vary over time. The second conclusion is that our first attempt at addressing possible violations of one of our identifying assumptions, namely the absence of anticipation due to additional information shocks, still leads to the same conclusion: workers use the information contained in the notification shock to anticipate their future treatment.

The third panel of Table 4 shows that the modeling of unobserved heterogeneity matters for the magnitude of the estimates, especially for the effect of notification on treatment, γ_P. However, qualitatively speaking, the results are similar in that evidence of

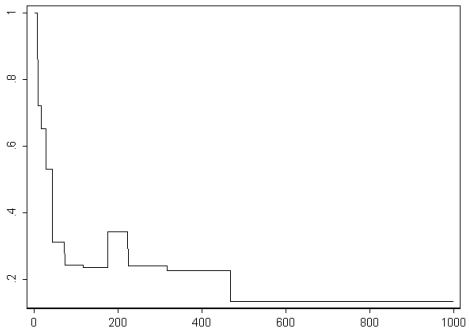
treatment anticipation remains strong when incorporating more groups of unobserved heterogeneity.

4.2 Hazard rates

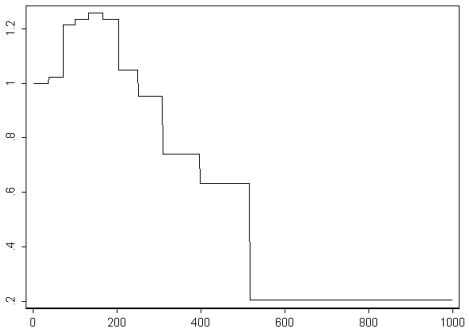
This subsection discusses results on time dependence, observed and unobserved heterogeneity in the three processes at play. All estimates are from the partial-information model using the benchmark specification. They thus correspond to the results in Table 3.

Time dependence. We start with the estimates of the λ functions, which, together with the treatment parameters, capture time dependence in model (9). Figure 1 shows displays these functions. Recall that the rates on the first interval are normalized, so these results are only qualitative. Recall 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.

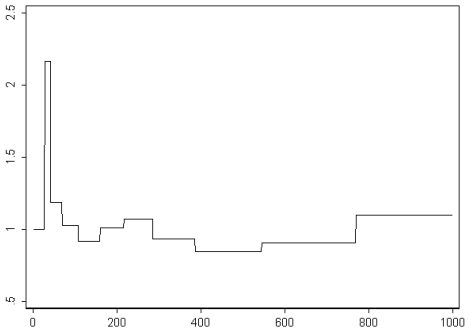
Figure 1: Time-dependent components of hazard rates (t in days, hazards $\times 10^4$)



$\lambda_P(t)$



$\lambda_Z(t)$



$\lambda_Y(t)$

Clearly, workers are more likely to receive notification upon entering unemployment, and also after 200 days. These results are consistent with the timing of interviews, as job seekers are obliged 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 decreases steadily and, after 550 days, 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).

Table 5: Effects of individual heterogeneity on hazard rates

Observed	β_P		β_Z		β_Y	
1{male}	-.249	(.010)***	-.172	(.013)***	.047	(.004)***
age	1.092	(.043)***	1.905	(.060)***	-1.308	(.014)***
age^2	-1.107	(.044)***	-2.003	(.061)***	1.170	(.014)***
experience	-.345	(.015)***	-.050	(.018)***	-.295	(.005)***
experience^2	.227	(.015)***	.050	(.019)***	.222	(.005)***
French	-.177	(.012)***	.378	(.018)***	-.220	(.005)***
children	-.036	(.013)***	.039	(.017)**	-.020	(.005)***
married	.097	(.012)***	.0163	(.016)	.017	(.004)***
blue collar	.280	(.024)***	-.388	(.035)***	.081	(.008)***
white collar unskilled	.437	(.020)***	-.373	(.026)***	.079	(.007)***
white collar skilled	.210	(.016)***	-.258	(.018)***	-.031	(.005)***
technical	.184	(.022)***	-.249	(.026)***	-.200	(.008)***
supervisor	.235	(.025)***	-.040	(.027)	-.049	(.009)***
junior high school drop out	.054	(.019)***	-.688	(.029)***	.212	(.007)***
junior high school degree	.232	(.020)***	.082	(.025)***	.120	(.007)***
high school drop out	.220	(.016)***	-.157	(.021)***	.140	(.006)***
high school degree	.207	(.017)***	.017	(.021)	.017	(.006)***
university drop out	.280	(.019)***	.179	(.023)***	-.007	(.007)
% treated last year	-5.637	(.196)***	.594	(.232)***	2.769	(.069)***
growth of unemp. inflow	-.258	(.037)***	.151	(.035)***	-.736	(.013)***
Unobserved	$\ln(V_P)$		$\ln(V_Z)$		$\ln(V_Y)$	
type 1	1.949	(.045)***	1.658	(.038)***	2.741	(.021)***
type 2	2.895	(.027)***	-1.492	(.044)***	3.638	(.010)***
Pr(type 1)	.289	(.008)***				

Individual heterogeneity. We now turn to the estimates of the effects of observed and unobserved individual heterogeneity on the three hazard rates. The first part of Table 4.2 shows the estimates of the β parameters, that is the effect of observed characteristics. Since we assumed a log-linear specification for the ϕ functions, a given characteristic 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).

We find that notification and training are primarily targeted at women and younger workers. 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. The same pattern can be observed for education, as high and junior high school drop outs are less likely to be trained than university graduates even though they are more likely to receive a notification.

The estimated parameters of the unobserved heterogeneity distribution suggest that, in terms of covariates that are excluded from X, two types of individuals can be distinguished. However, this finding is an artifact of the adoption of a discrete unobserved heterogeneity distribution, and the interpretation of the underlying types takes the assumed discrete functional form rather literally. Having said that, we note that "type 1" individuals in Table 4.2 are more likely to be notified but also have a high training rate per se. They may exhibit a high motivation for training, which may be correlated with the notification rate if the interviews with the caseworkers reveal this motivation (note that such an interpretation hinges on whether such factors are captured by X or not). The other type ("type 2") has a particularly low training rate per se. It is possible that a relatively large fraction of this type of individuals is ineligible for unemployment insurance. Recall that the latter information is not available in our data. Eligible job seekers benefit from improved access to funding of their training programs, as this funding is granted by the unemployment insurance system itself, and not by the state or the regions. In contrast to them, non-eligible individuals may be sufficiently motivated to be notified by the caseworkers but may not succeed in finding public funding for program entry.

4.3 Additional information shocks

We now consider the partial model described in subsection 3.2 and include two additional information shocks: a so-called A-shock, which is the first ANPE action realized after unemployment starts (other than a notification), *and* a so-called B-shock, which is the first ANPE action after notification has been received. Including these shocks into the analysis serves two purposes. First, it provides an additional robustness check for our main result as we will be able to verify whether the effects estimated above remain strong if we account for other information shocks. Secondly, the estimates offer additional insights into interactions between caseworkers and job seekers as well as insights into the way in which new information beyond notification affects worker behavior regarding training and/or exit from unemployment.

A-shocks are common: we observe an A-shock in 65% of the spells. This is not surprising, as job seekers are supposed to meet their caseworker at the beginning of the spell, and

this very first meeting can be the first A-shock. B-shocks are far less common: we only observe a B-shock in 4% of the spells. To some extent this is because by construction they can only be observed after P has been realized. Among spells that include a notification, B-shocks occur in 48% of the cases.

The estimates of the key parameters are shown in Table 6. First, note that all three information shocks seem to have a positive effect on training participation ($\gamma > 0$) and to trigger a response from the job seeker ($\delta \neq 0$). Notification has the strongest effects, whether it is preceded by another information shock (third row of Table 6) or not (second row). The point estimates have changed slightly with respect to the baseline model, with γ_P being lower and δ_P being larger, as large as the effect we found in the first month following notification (see Table 4). Still, the sign, significance and magnitude of γ_P and δ_P are in line with what we found without accounting for A and B.

Table 6: Effects of additional information shocks

	Effect on training			Effect on exit from unemployment		
A-shock	γ_A	0.359	$(0.018)^{***}$	δ_A	-0.246	$(0.004)^{***}$
P-shock	γ_P	2.171	$(0.020)^{***}$	δ_P	-0.611	$(0.010)^{***}$
P-shock, net of A	$\gamma_P - \gamma_A$	1.812	$(0.014)^{***}$	$\delta_P - \delta_A$	-0.365	$(0.010)^{***}$
B-shock, net of P	$\gamma_B - \gamma_P$	0.043	$(0.018)^{***}$	$\delta_B - \delta_P$	0.293	$(0.013)^{***}$

Note: A (resp. B) is the 1st shock before (resp. after) notification P.

The A-shock has qualitatively similar effects as notification: it increases the probability to be treated and lowers the probability to leave unemployment. These effects are however much smaller in absolute value than those of notification. This is expected as, in our data, notification is the ANPE action that kicks off the process towards training by putting the job seeker in contact with a training provider. It could be that some information regarding a future training program transpired during an early meeting between the job seeker and the caseworker. This information shock may have an impact on the probability to be treated or may affect the worker's job search strategy. We do not expect these effects to be large as the worker does not yet have the official information about his future treatment status. This is what we observe in Table 6 as γ_A is positive but much smaller than γ_P and δ_A is negative but much less than δ_P.

The first shock received after notification has a different effect. It increases the treat-

ment rate, but only marginally so in comparison to a notification. However, it has a significant and positive effect for notified workers, $\delta_B - \delta_P = 0.293$, on the rate of leaving unemployment. This suggests that post-notification ANPE actions create a threat effect. In sum, for those spells in which a B-shock is realized, it somewhat reduces the attraction effect of the notification, in the time interval from the B-shock until training.

Lastly, we should comment on the significance of the effects of the A- and B-shocks in Table 6. Formally, these imply that P is not the only relevant information shock. Comparing the estimates of γ_P and δ_P in Table 6 with those in Table 3, we see that omitting these additional information shocks generates a bias in the point estimation of the effects of notification, but the results remain qualitatively similar in that notification always has a positive impact on the training rate and a negative impact on the rate of leaving unemployment. Moreover, the effects are much stronger for notification than for the two other shocks (-1.13 and -0.66 against -0.088 and 0.13), which makes sense given the definition of notification used in our application.

4.4 The role of anticipation and information shocks in the evaluation of training programs

In this subsection, we examine how ignoring notification impacts the analysis of the effects of training programs on unemployment. To this end, we estimate two different models. The first model consists of two duration processes, duration until training Z and until exit from unemployment Y, thus completely ignoring notifications.[21] We refer to this as the "standard model". The second model, which has been referred to previously as the "full" model, includes all three processes: Z, Y and the duration until notification of treatment P. We estimate these two models separately by maximum likelihood allowing for $R = 2$ unobserved-heterogeneity types. The results are in Table 7.

We start with the top panel, showing the estimated effect of training on the log hazard rate of unemployment. We allowed for this effect to vary over time: δ_Z^0 (resp. δ_Z^4, resp. δ_Z^{12}) is the effect up to 4 months (resp. between 4 and 12 months, resp. more than 12 months) after the start of the program. In the first column, we have estimates from the "standard model" with just two processes Z and Y. We find that participating in a training program decreases the probability to leave unemployment in the first four months by 39% (≈1-exp(-0.497)) but significantly increases this probability after 4 months, by 71% (≈exp(0.538)-1), even more so after a year. This is the well-known locking-in effect (see Lechner et al., 2011), whereby individuals who are treated first experience a drop in their job finding rate (mostly due to the participation in the program).

[21] The hazard rates of Z and Y follow (9) where $\delta_P = \gamma_P = 0$.

Table 7: Effects of training programs and notification

	Model (Z, Y)		Model (P, Z, Y)	
Effect of training on unemployment exit				
- first 4 months (δ_Z^0)	-0.497	$(0.024)^{***}$	-0.633	$(0.021)^{***}$
- next 8 months (δ_Z^4)	0.538	$(0.018)^{***}$	0.412	$(0.016)^{***}$
- after a year (δ_Z^{12})	0.989	$(0.015)^{***}$	0.906	$(0.015)^{***}$
Effect of notification on unemployment exit (δ_P)	—		-0.334	$(0.010)^{***}$
Effect of notification on training (γ_P)	—		3.105	$(0.033)^{***}$

If we include notifications P and thus move on to column 2, the effects of training programs are qualitatively similar to those found with the first model. The estimate of δ_P indicates that the exit rate out of unemployment decreases after a notification (just like in subsection 4.1). Hence the δ_Z-estimates in the top panel and second column should be interpreted with respect to this reference point. This means that the locking-in effect of training is not as strong as it first seems since the hazard decreases by 26% (≈1-exp(-0.633+0.334)) for notified individuals.

We draw two conclusions from Table 7. First, from a methodological point of view, our results shed light on the importance of accounting for information shocks in the evaluation of training programs. If an individual received notification and is then treated, the change in his hazard rate at the start of the training program is not only due to the program itself but also to anticipation effects that took place between notification and treatment. Secondly, from a labor market policy design point of view, our results show that training programs are an event in a process that starts earlier, with earlier actions undertaken by caseworkers, and may induce workers to stay unemployed before the training program starts. Training programs have a positive effect on the long-term probability to leave unemployment. Hence an active labor market policy should consider the trade off between this long-term positive effect and the combination of an attraction effect prompted by notification and a short-term locking-in effect of training programs.

The timing of information shocks is thus important. Table 6 showed that notification and information shocks sent prior to notification (respectively P and A) have an attraction effect as opposed to a threat effect whereas the converse is true for information shocks sent after notification (B). This may motivate a close monitoring of workers who have received notification (i.e. intensification of B shocks) in order to reduce the pre-training attraction period. Designing an optimal timing for information shocks and training is beyond the

scope of the present paper. However we believe that our results motivates further analysis of this issue.

Another important feature shown in Table 7 pertains to the robustness of our results on anticipation effects. The reported estimates of the effects of notification on unemployment and training are close to those in the partial-information model (see Table 3). This illustrates the identification result presented in subsection 3.2 i.e. the fact that our results and conclusions on treatment anticipation from subsection 4.1 do not hinge on ad hoc aspects of the specification of the effect of training programs.

4.5 Expected unemployment durations for counterfactual notification and treatment dates

From estimated effects on hazard rates it is difficult to see what the order of magnitude is in terms of days in unemployment. We therefore use simulations to compute counterfactual expected unemployment durations. In particular, we compute expected durations for a range of counterfactual notification dates and a range of counterfactual dates of entering the training program. In this, we follow the approach of Eberwein, Ham and LaLonde (2002). As discussed by them, examining effects on expected durations has the additional benefit that those effects tend to be more robust to misspecifications than estimated hazard rate effects are.

For each observation i in the sample we perform the following procedure. First, for a range of given values of p and z, we compute the expected unemployment durations $E_{Y,V}\left[Y(z,p)|X_i,V\right]$. To this end, we compute the density of $Y(z,p)$ using the estimated full model where we do not stop following individuals after notification or treatment. Recall that our data follow individuals for 4 or 5 years, so we do not face a problem of predicting high out-of-sample durations. Secondly, for given p and z, we average each of these computed expected durations over i (that is, over the sample distribution of X).

Table 8 presents the results. Each column corresponds to a counterfactual treatment date z. The first row, $p = 0$, gives the average unemployment duration if workers are assigned to be notified directly upon entry into unemployment. Subsequent rows replicate this for notification times at one, two or three months after entry, and at one month before the starting date z of the training program. The last row in each column corresponds to $p = z$, which means that no notification is given before the training starts. A counterfactual that is not shown in the table is the one where p and z are infinitely large, meaning that nobody is ever notified or treated. The average unemployment duration in that case is 347 days.

Table 8: Counterfactuals - Expected unemployment durations.

	Sample mean of $E_{V,Y}[Y(z,p)\|X,V]$				
$p\backslash z$	30	60	90	180	360
0	289	297	307	336	375
30	281	290	299	328	365
60		281	290	317	354
90			284	311	346
150				302	336
180				298	331
330					314
360					312

Note: p is the assigned notification date,
z is the assigned treatment date.

A clear pattern emerges from Table 8: the earlier notification is given, the longer the unemployment spell. This does not contradict our estimation results. Specifically, as the estimated δ_P is negative, notifications induce an "attraction" effect. The magnitude of this effect is far from negligible: postponing notification by a month at the beginning of an unemployment spell decreases unemployment duration by at least a week on average.[22]

This analysis with counterfactual assignment dates can also be used to illustrate the bias when ignoring notifications in the assessment of training program effects. This goes back to the topics discussed in subsection 4.4. Consider the estimated "standard model" with no notification shocks, denoted as model (Z, Y), from that subsection. We can compute the expected unemployment duration for different training dates z in that model (formally $E_Y[Y(z)|X,V]$ averaged over X, V). It turns out that the expected duration equals 269 days if $z = 30$, 271 if $z = 60$, 276 if $z = 90$, 293 if training starts at six months, and 313 days if it starts after a year. Comparing these numbers with those in Table 8 reveals non-negligible differences. The standard model seems to under-predict ex-

[22] As an alternative approach to the use of expected unemployment durations to gauge the attraction effect of notifications, one may compare the expected duration in the estimated model with the expected duration in counterfactual situations defined by different parameter values. Specifically, one may take the estimated model, overrule the estimated value of δ_P by replacing it with the value zero, and subsequently compute the implied expected unemployment duration, averaging again over the sample distribution of X. The latter expectation can then be contrasted to the expectation implied by the full estimated model. However, such a crude comparison is hampered by the fact that many individuals in the sample are not notified. We may deal with this by conditioning on being notified before leaving unemployment. In that case the counterfactual expectation is about 73 days smaller. This number highlights the importance of the attraction effect of notifications but it should not be interpreted as a structural policy analysis since the underlying empirical reduced-form model is not specified in terms of economic decisions made by agents.

pected unemployment durations. This is consistent with the over-estimation of the effect of training on reemployment for the "standard model" in Table 7.

5 Conclusion

We study the role of notifications in the evaluation of training programs. This analysis is conducted first in the context of a theoretical job search model and then through a statistical dynamic evaluation framework, which we take to a unique administrative data set with detailed information on the exchange of information between caseworkers and job seekers in Paris.

We find strong and robust empirical evidence of workers responding to notifications of future training. More precisely, notifications have a positive effect on the probability of starting a training program and a negative effect on the probability to leave unemployment. These effects remain strong even when we account for additional information shocks, which are shown to also have an effect on both the treatment probability and on unemployment duration. The main methodological implication is that econometricians interested in the evaluation of training programs in a dynamic setting should endeavor to collect as much data as possible on private information shocks and model the duration processes driving those shocks together with the treatment and outcome durations.

As for policy implications, our paper shows that notification of future training programs actually decreases the exit rate from unemployment. This "attraction" effect goes against several other evaluations, which provide evidence of some "threat effect" of ALMPs. We can interpret this result in the light of the job search model derived in our paper. This model tells us that we would get a threat effect if training was valued negatively by workers (because of, say, a strong locking-in effect) and if notified workers could not avoid training. In the French institutional setting, job seekers rarely face sanctions if they do not participate in a training program so a threat effect is unlikely. According to the same model, an attraction effect is rather indicative of a positive effect of training on workers' search environment and of an increase in the training probability due to notification. Hence the locking-in effect estimated during the first months following the start of a training program can be compensated by the large increase in the probability to leave unemployment after four months.

In our application, the locking-in effect of training programs is preceded by anticipation effects taking place before the program starts. These two effects decrease the probability to leave unemployment around the treatment date. This indicates that an important criterion when designing active labor market policies is not only the timing of the treatment but also the timing of the information shocks. The job search model derived in this paper could provide a useful starting point to conduct a structural welfare analysis of the timing of notification and training,

The analysis essentially assumes that the notification is homogeneous in the sense that its effects do not depend on unobservables. This is a clear limitation, given that the unobservables affect most relevant outcomes in the model. It would be an interesting topic for further research to extend the model by making notification effects dependent on unobserved individual characteristics. Such an extension makes the model more intricate. To prevent that results are primarily driven by ad hoc model assumptions, it may be useful to gather additional information on the notification contents. It is an open question whether such information can be collected.

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 J. Heckman (2007): "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.

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.H. and G.J. van den Berg (2003b), "A Simple Procedure for the Evaluation of Treatment Effects on Duration Variables", Working paper, IZA, Bonn.

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.

Brinch, C.N. (2007): "Nonparametric Identification of the Mixed Hazards Model with Time-Varying Covariates", *Econometric Theory*, 23, 349-54.

Crépon, B., Ferracci, M. and D. Fougère (2012): "Training the Unemployed in France: How Does It Affect Unemployment Duration and Recurrence?" *Annals of Economics and Statistics*, 107/108, 175-200.

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.

Eberwein, Ham and LaLonde (2002): "Alternative Methods of Estimating Program Effects in Event History Models," *Labour Economics*, 9(2), 249-278.

Ferracci, M., Jolivet, G. and G.J. van den Berg (2014): "Evidence of Treatment Spillovers Within Markets," *Review of Economics and Statistics*, 96(5), 812-823.

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, 239-59.

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

Honoré, B.E. (1991): "Identification Results for Duration Models with Multiple Spells or Time-Varying Covariates," Working paper, Northwestern University, Evanston.

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 (2011): "Long-Run Effects of Public Sector Sponsored Training in West Germany," *Journal of the European Economic Association*, 9(4), 772-84.

McCall, B. (1996): "Unemployment Insurance Rules, Joblessness, and Part-Time Work," *Econometrica*, 64(3), 647-82.

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.

Van den Berg, G.J. (2001): "Duration Models: Specification, Identification and Multiple Durations," in: Heckman, J., and E. Leamer, eds., *Handbook of Econometrics*, vol. 5, ch. 55. Amsterdam: Elsevier Science, 3381-460.

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 (2014): "Policy Discontinuity and Duration Outcomes," Working paper, IZA.

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.