Conditional Treatment and Its Effect on Recidivism*

José R. Carvalho**
Herman J. Bierens***

Abstract

The objective of this paper is to evaluate the effect of the 1985 "Employment Services for Ex-Offenders" (ESEO) program on recidivism. Initially, the sample has been split randomly in a control group and a treatment group. However, the actual treatment (mainly being job related counseling) only takes place conditional on finding a job, and not having been arrested, for those selected in the treatment group. We use a multiple proportional hazard model with unobserved heterogeneity for job search and recidivism time which incorporates the conditional treatment effect. We find that the program helps to reduce criminal activity, contrary to the result of the previous analysis of this data set. This finding is important for crime prevention policy.

Keywords: Econometric Program Evaluation, Multiple Risks, Recidivism.

JEL Codes: C41, K42.

*Submitted in September 2005. Revised in March 2007. The financial support of CAPES Foundation, Brazil, and the helpful comments of Naércio Menezes and one referee on a previous version of this paper are gratefully acknowledged. This paper is based on the first essay of my Ph.D. Dissertation at Pennsylvania State University.

**Graduate Program in Economics, CAEN/Universidade Federal do Ceará. Avenida da Universidade, Fortaleza, CE – Brasil. CEP:60020-181. E-mail: josecarv@ufc.br

***Graduate Program in Economics, Pennsylvania State University, USA.

1. Introduction

The modern literature on program evaluation can be traced back to the seminal contributions of statisticians such as Rubin (1974) and Rosenbaum and Rubin (1983), and is mainly concerned with the estimation of treatment effects under both experimental and non-experimental setups. The evaluation problem has found great applicability within the economics profession and is, indeed, one of the areas that has attracted a huge interest both empirically and theoretically in the last decade.

Such interest can be appreciated in Heckman et al. (1999), Wooldridge (2002) and the extensive literature cited therein. As a first necessary step, the econometric evaluation of a program begins with a complete description of the probability model under investigation. This means a detailed enumeration of all relevant aspects of the program, especially the causal links between treatment and outcomes. In the program evaluation parlance, this step is known to establish the counterfactuals. In its simplest form, the stochastic content of a program can be described by a random vector $(y_0; y_1; w) \in \mathbb{R}^3$. Where $(y_0; y_1)$ is a vector of potential outcomes denoting the outcome without treatment and with treatment, respectively; and w is an indicator of treatment received, $w = 1$ or not received, $w = 0$.

As simple as it may appear, this framework can model situations such as the effect of schooling on wages, the effect of participation in a training program on labor market prospects, the effect of having a child on the divorce probability of couples, the effect of subsidies on small firms' survival and so on. Not surprisingly, generality almost always trades off with complexity and the evaluation problem has provided researchers with a source of challenges. As a matter of fact, the potential correlation between $(y_0; y_1)$ and w, the selection problem, has been an important problem in the field. In fact, the contributions of econometricians to handle that problem are distinctive.

Although the framework for evaluating the effects of programs cited above has been quite general, it is not possible to accommodate all possible situations that might appear. For example, an interesting situation occurs when the timing of the intervention is an important characteristic of the program. Such kind of "timing feature" must be framed within a duration analysis context. For our purposes, a program in a duration analysis context is defined as any program in which either the vector of outcomes $(y_0; y_1)$ or the time of receiving the treatment w or both are duration variables. To evaluate those type of programs, the econometrician faces the challenges posed by both fields of evaluation of programs and duration analysis.

We call programs with an outcome described by a duration variable, duration outcomes (DO), and call those with a stochastic time of realization of w, duration treatment (DT). A third possibility is a program having both features: we call them duration outcome and treatment (DOT).[1]

The seminal paper of Ham and LaLonde (1996) posed the key question of evaluating DO programs: even though random assignment to receive the treatment is performed, if the outcome is a sequence consisting of a duration variable followed by any variable that occurs conditional on that duration variable, selectivity issues will arise. In short, DO programs are open to dynamic selection problems even in a social experiment. Note that if no other outcome followed the first duration, there would be no problem if the assignment were random: to calculate the effect on the first duration, a simple difference between mean duration of controls and treatments is a consistent estimator of average treatment effects. Hence, the selection problem arises if any variable whose occurrence depends on the realization of the first duration is the object under evaluation, regardless of whether it is a duration variable or not. The following example should shed some light upon this situation.

Suppose a program offers training with the intent of both shortening the time unemployed and raising wages of unemployed people. People are randomly assigned to both control and treatment groups and everybody is unemployed at the beginning of the program at t = 0. The training lasts for $\widehat{t}$ and after a period of time t^* a random sample is collected. Hence, the random sample is collected after a period of time of $\widehat{t} + t^*$ from $t = 0$. There are four possible realizations: controls still unemployed, (CSU), controls employed, (CE), treatments still unemployed, (TSU), treatment employed, (TE). The interest is in measuring the effect of the program on the initial wages. Apparently, a nice strategy would be to estimate average treatment effects by calculating the difference between the sample mean wage of TE people and CE people. However, this estimator is very likely to be biased. The treatment could interact with some individual heterogeneity (for simplicity, assume it is a binary random variable, v) in a way that treated people with $v = 1$ are placed faster than those with $v = 0$. Also, assume that individuals with $v = 1$ would have a faster rate of employment and a higher wage compared to those with $v = 0$, even in the absence of the program. Thus, a sample of treated people collected at $t = \widehat{t} + t^*$, will have, on average, a much higher wage than the sample of controls will have. As a consequence, the average treatment effect is overestimated. The reason for this bias is clarified in Ham and LaLonde (1996): *[the outcome of interest] is missing for some individuals and wh*ether it is missing depe*nds on an individual's experimental status and unobservables ...* .

The key message from Ham and LaLonde (1996) paper[2] is that if one wants to

[1] As far as we know, there is no attempt to estimate such models. Hence, we omit this topic. However, we think that estimation of such models is an interesting topic for future research.

[2] It is worth mentioning that DO programs have been explored before in the econometric

analyze programs with duration outcomes, even within a social experiment setup, it is necessary to rely on non-experimental methods. As long as one recognizes that most labor market related programs are of this form, their message rises in importance. Those authors test how important dynamic selection is in practice, both in their seminal paper and in Eberwein et al. (1997). Both papers try to evaluate experimentally two DO programs, the National Supported Work Demonstration (NSWD) and Job Training Partnership Act (JTPA). They conclude by stating that the longer a program keeps an individual out of the labor market, the more important the issue of dynamic selectivity will be when trying to estimate treatment effects. For instance, the JTPA program presents no evidence of dynamic selection, whereas the NSWD showed an important selectivity effect.

Another kind of setup of even more recent interest is the econometric evaluation of DT programs. In DT programs the timing of occurrence of the treatment is an essential feature of the stochastic model. So, the duration component of the program is the stochastic time of realization of the treatment. Before we discuss DT programs, it is worth motivating it with an example. Consider that an individual is in a certain state at time t_0, say unemployment, and that, for the sake of simplicity, he/she can move only to a specific state, say employment. There is a chance of receiving training in a center any time in the near feature starting at t_0. Training, if received before employment, is instantaneously delivered. We are interested in measuring the effect of this training on the time of finding a job.[3] First, note that we are interested not only in the occurrence of an event, but also in the time of its occurrence. This is in contrast with the literature of program evaluation that is concerned only with the binary indicator w. Second, the question of dynamic selectivity remains an important obstacle to overcome.

The simplest form of a DT program model has two basic duration variables: $(T_{basic}; T_{treat})$ representing the basic state and timing of treatment. Early empirical applications include Lillard (1993) who estimates the effect of having a child on the duration of marriage. In that paper, the time throughout which a couple remains married is T_{basic} and the time to have the first baby is T_{treat}. His model allows the hazard rate for the duration of marriage to shift after the birth of a baby. Another application is contained in Lillard and Panis (1996). They try to estimate the impact of marriage dissolution on the longevity of people. A labor market application appears in van den Berg et al. (2004). The development of identification of a specific class of models to deal with DT programs is considered in the seminal paper of Abbring and van den Berg (2003). Their paper deals with the development of non-parametric identification of treatment effects in duration

literature, for instance, in Card and Sullivan (1988) and Meyer (1990). Those papers did not address the key dynamic selection problem, though.

[3]Clearly, training must occur before leaving unemployment, otherwise the question about the effects of treatment is pointless.

models. Without relying on the existence of a "true" control group,[4] Abbring and van den Berg (2003) proves that treatment effects are identifiable from both single and multiple duration data. As expected from the latter type of sample, conditions for identification are weaker.

Building on the above-mentioned literature, we develop and actually estimate an econometric model for a DT program which is able to address the following distinctive features:

1. (parametric) identification of the treatment effect;
2. existence of a "true" control group;
3. incorporation of recidivistic behavior;
4. handling of interval-censored observations.

Before we proceed to the model, we succinctly describe the main characteristics of the DT program to be evaluated.

2. The Employment Services for Ex-offenders Program

2.1 Antecedents

The Employment Services for Ex-offenders Program, henceforth ESEO program, originated from an agreement about the likely failures of past programs to re-employ ex-offenders in the USA. The Life Insurance for Ex-offenders (LIFE) and the Transitional Aid for Ex-offenders (TARP) are two early examples of employment services for ex-offenders. Both programs offered financial assistance as well as job placement services. The two programs reached similar conclusions: while financial assistance appeared to decrease the recidivism rate, job placement had little or no effect on reducing criminal activity, unless for those who succeeded in securing a job for a long time.

These early results should not be interpreted as a failure but, in fact, should be viewed as just a first step to the design of better programs. The lack of follow-up after placement was conjectured as the main obstacle to the complete success of such programs. As singled out by Milkman et al. (1985): *"Historically, employment services programs have severed contact with the client immediately after job placement. If any follow-up occurs, it is usually limited to periodic telephone contact with the employer to determine if the client is still employed. The programs generally cease to provide support ... virtually abandoning him [client] during this crucial time in his adjustment to life outside of the institution."*

The new paradigm of employment services for ex-offenders has resulted in the development of programs that had a strong preoccupation with the post-placement

[4] The same individual before the treatment serves as *his/her* own control.

of their clients. These programs have designed follow-up strategies to overcome the major criticism of past experience. From that perspective, the ESEO program[5] emerged as an important opportunity to assess the efficacy of employment services programs that contained a follow-up component.

2.2 Institutional framework

There are four important institutional aspects in any employment program: the eligibility rule, the assignment (between controls and treatment) scheme, provision of treatment and outcome measurement. The first two aspects will be dealt with in Section 3, where the details about the available data set are discussed. Hence, in the ESEO program,[6] after being assigned to either the control or treatment group the clients stepped into the intake unit, where they received initial guidance, screening and evaluation by an intake counselor. While still in this first phase, to secure survival up to the job search phase, the intake counselor offered minimal assistance services such as food, transportation, clothing, etc.

After intake, the client enters the second phase that will prepare him/her to develop job search skills: brief job development seminar which deals with issues such as appropriate dress and deportment, typical job rules, goal setting, interviewing techniques, and job hunt strategy. It is assumed that the time spent in the first and second phases are not random and negligible compared to the search phase and to the average duration of the outcome. The next and final phase of the provision of treatment is the job search assistance. This is the traditional job search assistance type of service, as described by Heckman et al. (1999). The job search assistance is the stage in the ESEO program that is offered equally to both controls and treatments. The difference begins with placement. Controls were not helped after placement, whereas treatments started receiving follow-up help just after the employment relationship started. Special follow-up services consisted basically of crisis intervention, counseling and, whenever necessary, reemployment assistance. These services lasted six months, and data from controls and treatments were collected at 30, 60 and 180 days after placement. For a more detailed exposition of each common service offered and received as well as those services specific to each individual program, one should refer to Timrots (1985), Milkman et al. (1985) and Milkman (2001).

[5] Actually, what is called the ESEO program is a set of three programs, *i.e.*, the Comprehensive Offender Resource System, in Boston, the Safer Foundation, in Chicago and Project JOVE, in San Diego. Of course, no program was identical with one another. However, specific attributes were not relevant so as to deserve separate analysis.

[6] We closely follow Milkman et al. (1985).

3. The ESEO Program Data Set

The ESEO data set consists of 2,045 individuals who participated in one of the three programs: 511 in Boston, 934 in Chicago and 600 in San Diego. However, the Inter-University Consortium for Political and Social Research (ICPSR) only made 1,074 usable observations available: 325 in Boston, 489 in Chicago and 260 in San Diego.[7] A large amount of information, sometimes very detailed, was collected from all sites. That can be broadly classified into three main categories:

- **Background variables**: demography, criminal history, employment history, educational achievement, and so on;
- **Program variables**: length of search, program participation record, reasons for dropout, placement features (wage, number of hours, match quality), and so on;
- **Outcome variables**: number of arrests, date of first arrest, self-reported arrests for placed people only, and so on.

A first important empirical issue is related to the characterization of the population being sampled. Unless very special assumptions are evoked, the validity of our findings can not be extrapolated beyond the population under sampling. In order to be eligible to participate in the ESEO program an individual must have the following background:[8]

1. Participants voluntarily accepted program services;
2. Participants had been incarcerated at an adult Federal, State, or local correctional facility for at least 3 months and had been released within 6 months from program participation;
3. Participants exhibited a pattern of income-producing offenses.

From the eligibility criteria it is clear that our population is a special, indeed very special, subset of the population of ex-offenders. Also, since participation is voluntary and there is no information on non-participants (those who did not choose to participate even though they fulfilled requirements 2 and 3), it is not possible to assess the potential bias on the sample induced by this selection scheme. Then, any result emerging from our econometric model must be interpreted considering

[7] The data set used in our present analysis comes from the Inter-University Consortium for Political and Social Research, henceforth ICPSR, under study number 8619. One may wonder about the likely bias arising from not using the whole sample. Unfortunately, we have no way to evaluate it, since ICPSR did not provide us with more information.

[8] For institutional details about the ESEO program, we closely followed the only 2 available published documents, *i.e.*, Milkman et al. (1985) and Timrots (1985).

those two initial issues. After this preliminary discussion, we should proceed to analyzing the available sample.

Given the initial sample, the individuals were randomly assigned to either the treatment or control group. Controls received the standard services, and treatments received, in addition to that, emotional support and advocacy during the follow-up period of 180 days after placement. Two durations are of great importance: time spent searching a job and recidivism time. However, these two variables are grouped.

The point of departure for the choice of the covariates is Schmidt and Witte (1988): age at release, time served for the sample sentence, sex, education, marital status, race, drug use, supervision status, and dummies that characterize the type of recidivism. However, we also pay close attention to the criminological literature on recidivism, for instance Gendreau et al. (1996).

The literature on unemployment (and job search) duration has been refined since the 1970s. Nowadays, it has a status of a complete theory of unemployment, as it appears in Pissarides (2000). Its empirical contents have been developed since the late 1970s and this first wave of empiricism is characterized by being concerned with "reduced" type models. A good account of this first phase can be found in Devine and Kiefer (1991). A final wave is characterized by advocating a "structural" approach to estimation and inference on such models. An updated account of that appears in van den Berg (1999). There also have been studies similar to ours that try to measure the effect of programs in a context of a model of unemployment and job search duration. For instance, Abbring et al. (2005), Eberwein et al. (1997) and van den Berg et al. (2004).

In view of those studies, a set of important covariates has been singled out. This set is composed basically of schooling, sex, age, and race. Together with the covariates related to recidivism, and the endogenous variables, the model variables are:

Endogenous variables

- **ATTRITION**: Indicator of attrition status. **ATTRITION** $= 1$ means the individual is either a "no show" or a "dropout", **ATTRITION** $= 0$ otherwise;

- **SEARCH**: Discrete variable indicating which interval[9] the search duration belongs to. **SEARCH** $= \{1, 2, 3 \textit{ or } 4\}$;

- **CRIME**: Discrete variable indicating which interval[10] recidivism belongs to. **CRIME** $= \{1, 2, 3 \textit{ or } 4\}$;

[9] See Table 1.
[10] See Table 1.

Table 1
Duration intervals

Number	Interval	Days
1	(0,30]	30
2	(30, 180]	150
3	(180, 360]	180

The search duration does not need any explanation, but the meaning of "recidivism" is not unambiguous. There are two ways to measure recidivism outcomes in the ESEO program: through count data or duration data. Detailed data on the number of arrests from date of release to the end of the program for all clients were gathered at the respective state police departments. Those were the data used in the original evaluation made by Milkman et al. (1985). Also, there are available data on the first arrest after release. The latter is what we will use as the duration of recidivism.[11] Thus, in the sequel "(duration of) recidivism" should be interpreted as the time between release and first rearrest.

Exogenous variables

- **GROUP**: Indicator of group participation. **GROUP** = 0 means control, **GROUP** = 1 means treatment;
- **DRUG**: Indicator of the use of drugs during the last 5 years. DRUG = 0 means no use, DRUG = 1 otherwise;
- **RACE**: Indicator of race. RACE = 0 means white, RACE = 1 means non-white;
- **SEX**: Indicator of sex. SEX = 0 means female, SEX = 1 means male;
- **EDUC**: Discrete variable describing educational attainment. EDUC = 0 if individual has from 2 to 8 years of schooling, EDUC = 1 if he/she has from 9 to 12 years or GED, and EDUC = 2 if he/she has more than 12 years of schooling;
- **AGE**: Age of ex-convict, in years;
- **SANDIEGO**: Indicator of city. SANDIEGO = 1 means San Diego, SANDIEGO = 0 means either Chicago or Boston;
- **CHICAGO**: Indicator of city. CHICAGO = 1 means Chicago, CHICAGO = 0 means either San Diego or Boston;

[11]Three possible definitions of recidivism are considered in the criminological literature: rearrest, reconviction and reincarceration. It seems that rearrest has been proven to be the most reliable among the three possible measures, as reported in Beck and Shipley (1989), and Maltz (1984).

- **AGEFIRST**: Age at first arrest, in years;
- **MOSUNEMP**: Number of months unemployed before baseline incarceration;
- **LASTWAGE**: Weekly wage before taxes in most recent job (before baseline incarceration).

Summaries of descriptive statistics of the covariates are given in Tables 2, 3 and 4. They correspond to the overall sample, subsample of controls ($GROUP = 0$) and subsample of treatments ($GROUP = 1$), respectively.

To have an idea about the empirical duration distribution, Tables 5 and 6 present the frequencies of the joint distributions for search time and recidivism time for controls and treatments with no attrition, respectively. More detailed information about the treatment group appears in Appendix A.

Table 2
Descriptive statistics

Variable	Observations	Min	Max	Mean	Stdev
ATTRITION	1074	0	1	0.5316	0.4992
GROUP	1074	0	1	0.5744	0.4946
DRUG	1074	0	1	0.3752	0.4844
RACE	1074	0	1	0.3352	0.4722
SEX	1074	0	1	0.8929	0.3093
EDUC	1074	0	2	1.0288	0.4466
AGE	1074	16	59	27.4497	6.5739
SANDIEGO	1074	0	1	0.2420	0.4285
CHICAGO	1074	0	1	0.4553	0.4982
AGEFIRST	1074	6	44	16.4543	4.5406
MONUNEMP	1074 (147 miss. obs.)	0	97	10.1467	19.5084
LASTWAGE	1074 (148 miss. obs.)	20	998	195.0119	114.7654

Table 3
Descriptive statistics – CONTROLS

Variable	Observations	Min	Max	Mean	Stdev
ATTRITION	457	0	1	0.7067	0.4557
DRUG	457	0	1	0.3435	0.4754
RACE	457	0	1	0.3172	0.4659
SEX	457	0	1	0.8927	0.3097
EDUC	457	0	2	0.9890	0.4514
AGE	457	16	53	26.9934	6.4319
SANDIEGO	457	0	1	0.1641	0.3707
CHICAGO	457	0	1	0.5229	0.5000
AGEFIRST	457	7	37	16.2800	4.0148
MONUNEMP	457 (73 miss. obs.)	0	97	10.7135	21.1003
LASTWAGE	457 (72 miss. obs.)	27	998	188.7506	106.8799

Table 4
Descriptive statistics – TREATMENTS

Variable	Observations	Min	Max	Mean	Stdev
ATTRITION	617	0	1	0.4019	0.4906
DRUG	617	0	1	0.3987	0.4900
RACE	617	0	1	0.3484	0.4768
SEX	617	0	1	0.8930	0.3093
EDUC	617	0	2	1.0583	0.4411
AGE	617	17	59	27.7876	6.6622
SANDIEGO	617	0	1	0.2998	0.4585
CHICAGO	617	0	1	0.4051	0.4913
AGEFIRST	617	6	44	16.5834	4.8930
MONUNEMP	617 (74 miss. obs.)	0	97	9.7458	18.3089
LASTWAGE	617 (76 miss. obs.)	20	998	199.4677	119.9605

4. An Econometric Model of the ESEO Program

4.1 Identification of treatment effect

To the best of our knowledge, Abbring and van den Berg (2003) is one of the first attempts to model treatment effects in a context of duration analysis that rigorously discusses nonparametric identification. However, the type of treatment effect identified by these authors is not the same treatment effect used in the literature on econometric program evaluation, because Abbring and van den Berg (2003) do not consider the presence of a control group as it is traditionally present in evaluation studies. Nevertheless, the treatment effect in our model is identified, but we have established this empirically rather than theoretically.

Table 5
Frequency distribution – CONTROLS

DURATION INTERVAL	SEARCH		CRIME	
	Frequency	Percent	Frequency	Percent
1	78	58.2	4	3
2	50	37.3	29	21.6
3	6	4.5	21	15.7
4	0	0	80	59.7
Total	134	100	134	100

Table 6
Frequency distribution – TREATMENT

DURATION INTERVAL	SEARCH		CRIME	
	Frequency	Percent	Frequency	Percent
1	185	50.1	15	4.1
2	151	40.9	79	21.4
3	28	7.6	68	18.4
4	5	1.4	207	56.1
Total	369	100	369	100

4.2 The Model

4.2.1 Absence of treatment

The latent dependent variables in our model are T_s, the job search time since release from prison, and T_c, the time of the first arrest after release from prison. Let $V \in \mathbb{R}_+$ be a random variable representing unobserved heterogeneity. In the absence of treatment the model could be specified according to the approach advocated by van den Berg (2000): conditional on the unobserved heterogeneity V and the exogenous variables in a vector X, the durations T_s and T_c are independent. Adopting a proportional representation for the hazard functions,

$$\theta_s(t|X,V) = \lambda_s(t) \cdot \phi_s(X) \cdot V \tag{1}$$

$$\theta_c(t|X,V) = \lambda_c(t) \cdot \phi_c(X) \cdot V \tag{2}$$

The conditional survival functions, given X and V, for each of the durations T_s, T_c are

$$S_s(t|X,V) = P(T_s \geqslant t|X,V) = \exp\left(-V \cdot \phi_s(X) \cdot \int_0^t \lambda_s(\tau)d\tau\right) \tag{3}$$

$$S_c(t|X,V) = P(T_c \geqslant t|X,V) = \exp\left(-V \cdot \phi_c(X) \cdot \int_0^t \lambda_c(\tau)d\tau\right) \tag{4}$$

Hence, the joint survival function conditional on X and V is:

$$\begin{aligned} & S(t_s,t_c|X,V) = P\left[T_s \geq t_s, T_c \geq t_c|X,V\right] \\ = \; & \exp\left(-V \cdot \left(\phi_s(X) \cdot \int_0^{t_s} \lambda_s(\tau)d\tau + \phi_c(X) \cdot \int_0^{t_c} \lambda_c(\tau)d\tau\right)\right) \end{aligned} \tag{5}$$

Finally, in order to tighten the durations T_s, T_c together and make them dependent conditional on X only, the random variable V has to be integrated out. Given a specification $G(v)$ of the distribution function of V, the joint survival function conditional on X alone is:

$$S(t_s,t_c|X) = \mathcal{L}\left(\phi_s(X) \cdot \int_0^{t_s} \lambda_s(\tau)d\tau + \phi_c(X) \cdot \int_0^{t_c} \lambda_c(\tau)d\tau\right)$$

where $\mathcal{L}(.)$ is the Laplace transform of G:

$$\mathcal{L}(s) = \int_0^{\infty} \exp\left(-v.s\right) dG(v), \;\; s \geq 0$$

4.2.2 Incorporating treatment

The key issue now is how to incorporate treatment in this framework. Let the dummy variable W represent group participation: $W = 1$ if the individual is selected in the treatment group, and $W = 0$ if selected in the control group. Then treatment is received if

1. The individual is selected in the treatment group: $W = 1$.
2. The job search has ended before the first arrest: $T_s < T_c$.

The problem is now that due to the latter condition it is impossible to build the effect of treatment directly into the joint survival function (5) without sacrificing the conditional independence of T_s and T_c given X and V. However, note that without assuming conditional independence we can still factorize out the joint density of T_s and T_c conditional on X, V, and W, as a product of conditional densities, say:

$$\begin{aligned} & f\left(t_s, t_c | X, V, W\right) \\ = \; & f_c\left(t_c | T_s = t_s, X, V, W\right) . f_s\left(t_s | X, V, W\right) \end{aligned}$$

Consequently, the corresponding joint survival function can be written as

$$\begin{aligned} & S(t_s, t_c | X, V, W) \\ = \; & P\left[T_c \geq t_c, T_s \geq t_s | X, V, W\right] \\ = \; & \int_{t_s}^{\infty} \int_{t_c}^{\infty} f_c\left(\tau_c | T_s = t_s, X, V, W\right) d\tau_c \; f_s\left(\tau_s | X, V, W\right) d\tau_s \end{aligned}$$

Therefore, in modeling the joint survival function of T_s and T_c conditional on X, V, and W we can still use a similar setup as before, as follows.

First, model the conditional hazard function of T_c conditional on $T_s = t_s, X, V, W$ as

$$\begin{aligned} & \theta_c(t_c | t_s, X, V, W) \\ = \; & [(1-W)\phi_c(X) + W.(1 - I(t > t_s))\phi_c(X) + W.I(t > t_s)\phi_c^*(X)] \\ & \times \lambda_c(t_c) \cdot V. \\ = \; & [\phi_c(X) + W.I(t > t_s)\left(\phi_c^*(X) - \phi_c(X)\right)] \times \lambda_c(t_c) \cdot V. \end{aligned}$$

where $I(.)$ is the indicator function. If $W = 0$ this specification corresponds to the previous one in (2), but for $W = 1$ the effect of the treatment on recidivism is now incorporated:

$$\theta_c(t_c|t_s,X,V,W=1)=\begin{cases}\phi_c^*(X)\cdot\lambda_c(t_c)\cdot V & \text{if } t_s<t_c\\ \phi_c(X)\cdot\lambda_c(t_c)\cdot V & \text{if not}\end{cases}$$

where $\phi_c(X)$ is the same as in (2), and $\phi_c^*(X)$ is the systematic hazard during treatment. The corresponding conditional survival function of T_c is now

$$\begin{aligned}S_c(t_c|t_s,X,V,W) &= P(T_c\geqslant t_c|T_s=t_s,X,V,W)\\ &= \exp\left(-V\cdot\Lambda_c(t_c|t_s,X,W)\right)\end{aligned}$$

where

$$\begin{aligned}&\Lambda_c(t_c|t_s,X,W) \qquad (6)\\ =\;&\phi_c(X)\int_0^{t_c}\lambda_c(\tau)d\tau+W.\left(\phi_c^*(X)-\phi_c(X)\right)\int_0^{t_c}I(\tau>t_s)\lambda_c(\tau)d\tau\\ =\;&\phi_c(X)\int_0^{t_c}\lambda_c(\tau)d\tau+W.\left(\phi_c^*(X)-\phi_c(X)\right)I(t_c>t_s)\int_{t_s}^{t_c}\lambda_c(\tau)d\tau\end{aligned}$$

is the corresponding integrated hazard.

The conditional survival function of T_s is the same as before:

$$\begin{aligned}S_s(t_s|X,V,W) &= P(T_s\geqslant t_s|X,V,W)\\ &= P(T_s\geqslant t_s|X,V)=\exp\left(-V\cdot\Lambda_s(t_s|X)\right)\end{aligned}$$

where

$$\Lambda_s(t_s|X)=\phi_s(X)\int_0^{t_s}\lambda_s(\tau)d\tau$$

is the integrated hazard. Thus, the joint survival function of T_s, T_c conditional on X, V, and W is:

$$\begin{aligned}&S(t_s,t_c|X,V,W) \qquad (7)\\ =\;&P\left[T_c\geq t_c,T_s\geq t_s|X,V,W\right]\\ =\;&\int_{t_s}^{\infty}S_c(t_c|\tau,X,V,W)f_s\left(\tau|X,V,W\right)d\tau\\ =\;&\int_{t_s}^{\infty}\exp\left[-V\cdot\Lambda_c(t_c|\tau,X,W)\right]V\exp\left(-V\cdot\Lambda_s(\tau|X)\right)\phi_s(X)\lambda_s(\tau)d\tau\\ =\;&V.\phi_s(X)\int_{t_s}^{\infty}\exp\left[-V\cdot\left(\Lambda_c(t_c|\tau,X,W)+\Lambda_s(\tau|X)\right)\right]\lambda_s(\tau)d\tau\end{aligned}$$

where the last two equalities follow from

$$\begin{aligned} f_s\left(t|X,V,W\right) &= -\frac{\partial}{\partial t}S_s(t|X,V,W) \\ &= -\frac{\partial}{\partial t}\exp\left(-V\cdot\Lambda_s(t|X)\right) \\ &= V\exp\left(-V\cdot\Lambda_s(t|X)\right)\phi_s(X)\cdot\lambda_s(t) \end{aligned} \quad (8)$$

4.2.3 Baseline hazards

The baseline hazards $\lambda_s(t)$ and $\lambda_c(t)$ are assumed to have a Weibull specification:

$$\begin{aligned} \lambda_s(t) &= \lambda_s t^{\lambda_s-1},\ \lambda_s>0 \\ \lambda_c(t) &= \lambda_c t^{\lambda_c-1},\ \lambda_c>0 \end{aligned} \quad (9)$$

For search or unemployment durations, the Weibull hazards are flexible enough to capture any pattern of monotonic dependence typical of labor markets. See for example van den Berg et al. (1994). Regarding criminal behavior, "parabola" type hazards might be more appropriate. Generally, after release, the ex-criminal has a period of low criminal activity followed by a high criminal one. However, as long as the abscissa of the point of maximum of those parabolic hazards is close enough to the origin, the Weibull hazards are still a reasonable approximation. Then the integrated conditional hazards become

$$\begin{aligned} &\Lambda_c(t_c|t_s,X,W=1) \\ =\ &\phi_c(X)t_c^{\lambda_c}+\left(\phi_c^*(X)-\phi_c(X)\right)I(t_c>t_s)\left(t_c^{\lambda_c}-t_s^{\lambda_c}\right) \end{aligned} \quad (10)$$

and

$$\Lambda_s(t|X)=\phi_s(X)\cdot t^{\lambda_s} \quad (11)$$

Hence, the joint conditional survival function for the treatment group takes the form:

$$\begin{aligned}
& S(t_s, t_c|X, V, W = 1) \qquad (12)\\
= \quad & I(t_c > t_s)V.\phi_s(X)\exp(-V \cdot \phi_c(X)t_c^{\lambda_c}t)\\
\times \quad & \int_{t_s}^{t_c} \exp[-V \cdot (\phi_c^*(X) - \phi_c(X))\left(t_c^{\lambda_c} - \tau^{\lambda_c}\right)]\\
\times \quad & \exp[-V \cdot (\phi_s(X) \cdot \tau^{\lambda_s})]\lambda_s \tau^{\lambda_s - 1} d\tau\\
+ \quad & I(t_c > t_s)\exp[-V \cdot (\phi_c(X)t_c^{\lambda_c} + \phi_s(X) \cdot t_c^{\lambda_s})]\\
+ \quad & I(t_c \leq t_s)\exp[-V \cdot (\phi_c(X)t_c^{\lambda_c} + \phi_s(X) \cdot {t_s}^{\lambda_s})]
\end{aligned}$$

(see Appendix B for the details about the derivations involved), whereas for the control group:

$$\begin{aligned}
& S(t_s, t_c|X, V, W = 0) \qquad (13)\\
= \quad & \exp\left[-V \cdot \left(\phi_c(X) \cdot t_c^{\lambda_c} + \phi_s(X) \cdot t_s^{\lambda_s}\right)\right]
\end{aligned}$$

4.2.4 Systematic hazards

For the systematic hazards $\phi_s(X)$ and $\phi_c(X)$ we adopt the usual exponential specification:

$$\begin{aligned}
\phi_s(X) &= \exp(\beta_s' X) \qquad (14)\\
\phi_c(X) &= \exp(\beta_c' X)
\end{aligned}$$

where X contains 1 for the constant term. As to the specific hazard upon treatment, we assume that

$$\phi_c^*(X) = \delta\phi_c(X) = \delta\exp(\beta_c' X), \;\; \delta > 0$$

In the sequel, however, we will continue to use the notations $\phi_s(X)$, $\phi_c(X)$ and $\phi_c^*(X)$.

The parameter δ is the key parameter in our model, as it measures the effect of the ESEO program on the recidivism behavior of its participants. The parameter δ either inflates or deflates the systematic hazard function of recidivism upon placement. Its interpretation is:

- If $\delta > 1$, the program has a negative impact on recidivism, as it inflates the hazards for recidivism, therefore shortening the time between release and first arrest;
- If $\delta = 1$, the program has no effect;

- If $\delta < 1$, the program has a positive impact on recidivism, as it deflates the hazards for recidivism, therefore lengthening the time between release and first arrest.

4.2.5 Unobserved heterogeneity

The traditional[12] choice of the distribution of the heterogeneity variable V is the Gamma distribution, because its Laplace transform has a closed form expression: If $V \sim Gamma(\alpha, \zeta)$ then the Laplace transform of V is:

$$\mathcal{L}(s) = E\left[\exp(-s.V)\right] = (1 + s \cdot \zeta)^{-\alpha} \tag{15}$$

with derivative

$$\mathcal{L}'(s) = -E\left[V \exp(-s.V)\right] = -\alpha\zeta\,(1 + s \cdot \zeta)^{-\alpha-1} \tag{16}$$

Adopting the specification, it follows from (12) through (16) that:

$$\begin{aligned} & S(t_s, t_c|X, W = 1) \qquad (17) \\ = \quad & I(t_c > t_s)\alpha\zeta\phi_s(X) \\ \times \quad & \int_{t_s}^{t_c} [1 + \zeta(\phi_c(X)t_c^{\lambda_c} + (\phi_c^*(X) - \phi_c(X)t)(t_c^{\lambda_c} - \tau^{\lambda_c}) + \phi_s(X) \cdot \tau^{\lambda_s})]^{-(\alpha+1)} \\ \times \quad & \lambda_s \tau^{\lambda_s - 1} d\tau \\ + \quad & I(t_c > t_s)[1 + \zeta(\phi_c(X)t_c^{\lambda_c} + \phi_s(X) \cdot t_c^{\lambda_s})]^{-\alpha} \\ + \quad & I(t_c \leq t_s)[1 + \zeta(\phi_c(X)t_c^{\lambda_c} + \phi_s(X) \cdot {t_s}^{\lambda_s})]^{-\alpha} \end{aligned}$$

and

$$S(t_s, t_c|X, W = 0) = \left[1 + \zeta\left(\phi_c(X) \cdot t_c^{\lambda_c} + \phi_s(X) \cdot t_s^{\lambda_s}\right)\right]^{-\alpha} \tag{18}$$

Note that ζ cannot be identified. To see this, substitute (14) in (18):

$$\begin{aligned} & S(t_s, t_c|X, W = 0) \\ = \quad & \left[1 + \exp(\ln(\zeta) + \beta_c' X) \cdot t_c^{\lambda_c} + \exp(\ln(\zeta) + \beta_s' X) \cdot t_s^{\lambda_s}\right]^{-\alpha} \end{aligned}$$

Since $\beta_s' X$ and $\beta_c' X$ contain constant terms, $\ln(\zeta)$ can be absorbed into the constants. **Consequently, we will set** $\zeta = 1$.

[12] See, for example, Lancaster (1990).

4.2.6 Attrition

There are two types of attrition in our sample, namely "no show" if an individual does not participate at all in the job search stage of the program, and "quitting" of an individual during the job search stage. As to attrition, we decided to take a very pragmatic approach. Instead of modeling these two types of attrition jointly with job search and recidivism, we assume that the survival functions (17) and (18) apply **conditionally** on the absence of attrition, where attrition now includes "no show" and "quitting".

If an individual quits after finding a job, and this individual is in the treatment group, we will assume that the treatment effect is the same as for an individual who completes the treatment.

Let $I_A = 1$ indicate attrition, and $I_A = 0$ absence of attrition. We will specify the probability of attrition as a Logit model:

$$P[I_A = 1|X, W = w] = \frac{1}{1 + \exp\left(\gamma_j' X\right)},\ w = 0, 1 \qquad (19)$$

The parameters γ_j may be different for $w = 0, 1$ (controls and treatments, respectively).

4.2.7 Censoring

The actual durations T_s and T_c are not directly observed, but are only known to belong to particular intervals, *i.e.*, T_s and T_c are known to belong to one of the following four intervals appearing in Table 1.

There are 12 combinations where T_s and T_c are in different intervals: $T_s \in [a_i, b_i), T_c \in [c_i, d_i)$, say, where either $b_i \leq c_i$ or $d_i \leq a_i$. The remaining four cases, $T_s \in [a_i, b_i), T_c \in [a_i, b_i)$, will be treated as "other", because there are relatively few observations for which the latter applies, and secondly, the computation of $P\left(T_s \in [a_i, b_i), T_c \in [a_i, b_i)\right)$ is more complicated than in the non-overlapping cases.

Probabilities of the type $P(T_s \in [a, b), T_c \in [c, d))$ can be easily computed on the basis of the joint survival functions:

$$\begin{aligned} P(T_s \in [a, b), T_c \in [c, d)|X, W) &= S(a, b|X, W) - S(b, c|X, W) \\ &- S(a, d|X, W) + S(b, d|X, W) \end{aligned} \qquad (20)$$

4.3 The likelihood function

Let $I_i = [a_i, b_i) \times [c_i, d_i)$, $i = 1, ..., k$, be **disjoint** intervals in $\mathbb{R}^2_+$. For each individual j, assign a dummy variable $D_{i,j}$ such that $D_{i,j} = 1$ if $(T_{c,j}, T_{s,j}) \in I_i$, and let $D_{0,j} = 1 - \sum_{i=1}^{n} D_{i,j}$. Then for $i = 1, ..., k$,

$$\begin{aligned} P[D_{i,j} &= 1|X_j, W_j] = P\left[(T_{c,j}, T_{s,j}) \in I_i|X_j, W_j\right] \\ &= S(a_i, b_i|X_j, W_j) - S(b_i, c_i|X_j, W_j) \\ &- S(a_i, d_i|X_j, W_j) + S(b_i, d_i|X_j, W_j) \\ &= p_{i,j}(\theta) \end{aligned}$$

say, where

$$\theta = (\beta_s', \lambda_s, \beta_c', \lambda_c, \delta, \alpha)'$$

with $W_j = 0$ if individual j belongs to the control group, and $W_j = 1$ if he/she belongs to the treatment group. Moreover, the probability of an individual belonging to the category "other" is:

$$P\left[D_{0,j} = 1|X_j, W_j\right] = 1 - \sum_{i=1}^{k} p_{i,j}(\beta) = p_{0,j}(\beta)$$

Next, from equation (19) define

$$P\left[I_{A_j} = 1|X_j, W_j = w\right] = q_j(\gamma_w),\ w = 0, 1$$

Moreover, recall that we have assumed that

$$P\left(T_{c,j} > a, T_{s,j} > b|X_j, W_j, I_{A_j} = 0\right) = S(a, b|X_j, W_j)$$

Then the log-likelihood takes the form:

$$\begin{aligned}
& \log \mathbf{L}(\theta, \gamma_0, \gamma_1) \\
= & \sum_{j=1}^{n} I_{A_j}((1 - W_j) \ln q_j(\gamma_0) + W_j \ln q_j(\gamma_1)) \\
+ & \sum_{j=1}^{n} (1 - I_{A_j})[\sum_{i=0}^{k} D_{i,j} \ln p_{i,j}(\theta) + (1 - W_j) \ln(1 - q_j(\gamma_0)) \\
+ & W_j \ln(1 - q_j(\gamma_1))] \\
= & \sum_{j=1}^{n} I_{A_j}((1 - W_j) \ln q_j(\gamma_0) + W_j \ln q_j(\gamma_1)) \\
+ & \sum_{j=1}^{n} (1 - I_{A_j})[(1 - W_j) \ln(1 - q_j(\gamma_0)) + W_j \ln(1 - q_j(\gamma_1))] \\
+ & \sum_{j=1}^{n} (1 - I_{A_j}) \sum_{i=0}^{k} D_{i,j} \ln p_{i,j}(\theta) \\
= & \log \mathbf{L}_0(\gamma_0) + \log \mathbf{L}_1(\gamma_1) + \log \mathbf{L}_2(\theta)
\end{aligned}$$

where n is the sample size.

4.4 Estimation and inference

All econometric work (data manipulation, estimation and inference) was conducted by means of the econometric **EasyReg International**[13] package.

4.4.1 Attrition

The results for the logit estimation appear in Table 7 and Table 8 for controls and treatments, respectively. The first set of estimates are the estimates of the components of γ_0, and the second set corresponds to γ_1 The majority of the estimated regressors are not significant at the 10% level.[14]

[13]This freeware package was developed by the second author and can be downloaded from: http://econ.la.psu.edu/∼hbierens/easyreg.htm.

[14]For the sake of convenience, from now on, any reference made about parameter significance implicitly assumes a level of 10%.

Table 7
Controls

Parameters	ML estimate	t-values
Age	0.0369	1.59
First arrest	-0.0324	-0.97
Drug	0.3146	1.16
Race	-0.0427	-0.15
Sex	1.3071	3.13
Education	-0.1899	-0.70
Chicago	2.1590	6.97
Sandiego	1.2119	3.42
Monusemp	-0.0024	-0.38
Lastwage	-0.0018	-1.47
Intercept	-1.5860	-1.70
Log-Likelihood		-199.23
n		374

Table 8
Treatments

Parameters	ML estimate	t-values
Age	-0.0055	-0.35
First arrest	0.0454	2.21
Drug	0.0091	0.04
Race	0.2889	1.39
Sex	0.1707	0.51
Education	-0.1765	-0.81
Chicago	0.4660	1.92
Sandiego	0.2755	1.13
Mosunemp	0.0072	1.46
Lastwage	-0.0002	-0.35
Intercept	-1.5072	-2.43
Log-Likelihood		-340.21
n		526

For controls, only the variables Sex, Chicago and Sandiego are significant. A man has a higher probability of attrition than a woman. Belonging to the program located in Chicago, as well as in San Diego, raises the probability of attrition. Most of the components of γ_1 are insignificant, except First arrest and the dummy Chicago: for ex-inmates having their first arrest earlier, the probability of attrition is lower, and having served their last term in a Chicago prison increases the probability of attrition.

Interestingly, the estimated parameters for variables Monusemp and Lastwage suggest that employment history has no effect on attrition for both groups.[15] However, the available original variables related to employment history are not

[15] We thank an anonymous referee for drawing our attention to the fact that past employment and criminal history might be important determinants for attrition.

very good measures of past experience in the labor market. Such facts may well have an important role in the results obtained. With regard to criminal history, the ICPSR study assembled only the records after the prisoner was released. This fact precludes the use of these variables as predictor of recidivistic behavior in our setup.

4.4.2 Job search, recidivism, and treatment

Results appear in Table 9. In order to interpret the results, note that if a coefficient is positive and the corresponding X variable increases, then the whole hazard function will be inflated, hence the integrated hazard will be reduced, and so will the survival probability. Thus, failure will occur earlier. In the case of search duration, this implies that the average time of search (unemployment) will be lower, the higher the value of the X variable is. For the crime (recidivism) duration this implies that the expected time between release and rearrest will be reduced.

Some parameter estimation results for search duration appear to contradict well established facts in the literature on empirical search models. However, given the specific nature of our data (ex-criminals), there are some reasonable explanations for that. The demand side of the job market appears to be driven much more by the possibility that the future worker could commit a crime after being hired than by pure efficiency considerations. Also, the job market for ex-criminals is characterized by being of bad quality and by offering low wages. Such empirical evidence concerning job search for ex-inmates looks promising as a topic for future development.

The estimated parameter for age is negative and therefore, as expected, the job search time is higher the older the ex-inmate is. Males appear to have search time greater than that search time of women. This is the first result that contradicts empirical findings in search models. Indeed, the sex effect is significant. However, the male/female ratio of inmates is much higher than the 50% ratio out of prisons, so that males may present a higher potential threat of comitting a crime while employed. The positive coefficient of education means that more educated ex-inmates will find jobs faster than less educated ones, which is in accordance with the empirical search literature. See, *e.g.*, Devine and Kiefer (1991). The significantly negative coefficients of the dummies for Chicago and San Diego, indicate that job search time in Chicago and San Diego[16] is larger than in Boston, *ceteris paribus.* Finally, the parameter of the baseline hazards presents a rather surprising result. As shown in Lancaster (1990), a value of $\lambda_s = 1.582$ means that the search time presents positive dependence, or in other words, the longer an individual keeps searching the higher the probability of him/her finding a job. This is exactly the

[16] Of course, it is not possible to distinguished between local labor market conditions and program's features.

opposite of a lot of evidence found in studies of search in the labor market. For instance, see Devine and Kiefer (1991). A plausible explanation for the high value of λ_s (which is significantly greater than 1) is that the ex-inmate will have to cope with the difficulty in fiding a reasonable job soon after her release from prison, given the strong stigma prevalent in the society. Hence, sooner or later he/she will have no option but to accept any (low-paying) job that becomes available.

Table 9
Parameters estimate

Duration	Parameters	ML estimate	t-values
search	age	-0.075	-3.423
	first arrest	0.007	0.306
	drug	0.090	0.308
	race	0.167	0.583
	sex	-1.147	-2.302
	education	0.348	1.968
	chicago	-2.021	-5.540
	sandiego	-1.071	-3.021
	intercept	-2.600	-2.529
	λ_s	1.582	9.862
crime	age	-0.117	-5.320
	first arrest	-0.031	-1.319
	drug	-0.500	-1.773
	race	0.523	1.895
	sex	-1.670	-4.078
	education	0.200	1.100
	chicago	-0.361	-1.130
	sandiego	-0.026	-0.086
	intercept	-2.986	-3.147
	λ_c	1.676	10.179
heterogeneity	α	1.213	3.075
effect of treatment	δ	0.631	3.947
	Log-Likelihood		-6425.281
	n		1074

The impact of age on the expected recidivism time is significantly negative. This is in accordance with other studies in recidivism such as Schmidt and Witte (1988). Hence, *ceteris paribus*, an older ex-inmate will postpone his/her next crime. The estimated parameter of the dummy variable race (1 = non-white) is positive, but only borderline significant. Hence, non-whites seem to recidivate earlier than whites, which is in accordance with other empirical studies on recidivism, such as Schmidt and Witte (1988). The strongly significant negative value of the coeffient of sex appears to contradict the literature on criminal recidivism: females will commit a crime earlier than males. However, the sample consists only for about 11% of females, so that a very few bad ones among them may cause this effect. The city dummies do not have a significant effect. Also for recidivism the parameter λ_c is significantly greater than 1, which implies that the longer an ex-inmate is out without committing a crime the higher the probability of committing a crime in the future.

The parameter α is a nuisance parameter with no particular interesting interpretation other than that it is the expected value of the unobserved heterogeneity variable V.

The parameter δ is the key parameter on our econometric model. The estimated value is significantly less than 1. Hence, the program is **effective** as it increases the time between release and rearrest. This result stands in contrast with the original study of the ESEO program, as shown in Milkman et al. (1985).

In order to gain more perspective about the differences between our results and those from Milkman et al. (1985), henceforth "original paper", it is important to delve into more details about the latter. In the original paper a set of linear regression models is estimated. A common set of independent variables (age, sex, race, a dummy for treatment status, work history, criminal history and so on) is used to explain three separate models according to the dependent variable employed:

1. total number of arrests;

2. total number of income-producing arrests and;

3. total number of Part 1 arrests, *i.e.*, violent crime.

Their methodology to evaluate the "impact" of the ESEO program was based entirely on the analysis of the estimated parameter of the dummy for treatment status. The main conclusions were:

1. Some independent variables are significant predictors of recidivism behavior as defined, such as race, sex, marital status, criminal history, past drug use. Other variables do not significantly predict recidivism such as work history and dummy for location;

2. The dummy for treatment status is not significantly different from zero. From this result the authors concluded that *"The results of this evaluative research reveal that specialized, intensive assistance to ex-offenders during the initial months of their employment does not, as had been hypothesized, result in lower rates of long-term criminal recidivism"*.

The authors, as a logical step after the estimation exercise, pointed out that *"... there appears to be no justification for encouraging employment service programs to provide such pos-employment assistance"*. Apart from the obvious differences between the two approaches concerning recidivism definition, set of independent variables, functional form and so on, we believe that the main drawback of the original paper is that they completely disregard the timing effect of the treatment. This is evident by noting that all effect of the treatment is modeled by introducing the dummy for treatment status.[17] Hence, we believe that our results, by incorporating the timing of treatment, is a superior approach on that context. Next section concludes the paper and points out some avenues for future research.

5. Conclusion

By modeling the ESEO program as a mixed multivariate proportional hazards model,[18] where treatment is conditional on placement, we have merged two important fields of modern econometrics: survival analysis and econometric evaluation of programs. As far as we know, our paper is the first one to build this type of model and estimate it. The following paragraphs conclude by discussing the main achievements of the present paper, as well as by offering some possible ideas for future research.

First, our contribution has to do with the available data set. Even though this data set has been used before, it was restricted to the community of sociologists and criminologists. Despite the fact that search models have been estimated since the early 1980's, search by ex-inmates who participate in a program of reemployment is a novelty for the econometric audience. The estimated parameters appearing in Table 9, and the discussions that followed it show that some regressors have very different effects when compared to the traditional search model. Nonetheless, our available data presents some limitations. The main limitation of our data set is that it is grouped and this definitely imposes constraints on what can be identified from the model and makes our results less convincing. A good standard to be followed by criminologists and sociologist would be the methodology used by the agencies that collect unemployment data in the USA. Better data help a lot, specially in econometric evaluation of programs, as shown by Heckman et al. (1999).

Second, we have shown some evidence of how the process of search for jobs could be heavily influenced by the demand side of the market. More specifically, it would not be surprisingly that information asymmetries play a crucial role in this specific labor market. It is very likely that all prospective employers know that each of application comes from an ex-inmate, however knowledge of the past criminal history of each ex-convicts does not need to follow. Indeed, legislation regarding disclosure of criminal past records varies a lot within the USA. Hence, an interesting topic for future research would be the estimation of models that explicitly consider the information asymmetries existent in this market. We think this should be a nice starting point to address the actual debate about disclosure of criminal records and to evaluate its policy implications.

[17] Although it is a tricky task figuring out what people intentions are, the fact that people are randomized in the beginning of the program might have driven the authors to think that they have a "natural experiment" at hand. Our discussion on Section 1, hopefully, makes a point against this subtlety.

[18] See, Abbring and van den Berg (2000) for this nomenclature.

Third, the blending of survival analysis and econometric program evaluation represents our key contribution. We have set a model where the timing of treatment is explicitly considered. This stands in contrast with any other past study of econometric evaluation of programs. In fact, we are able to build an estimable model and estimate it. The estimated parameters clearly show that the timing of treatment is an important feature of social programs well neglected in the past. Nonetheless these initial accomplishments, there is still important topics for future development.

A first important question is the non- or semi-parametric estimation of hazard functions models with unobserved heterogeneity. We have used the simplest way of incorporating unobserved heterogeneity, *i.e.*, a fully parametrized Gamma random variable. Given our already complex model that nested two latent duration variables together with a treatment effect, the whole parametric setup was nothing but pragmatism. However, since the seminal paper of Heckman and Singer (1984), the high sensitiveness of the estimates to the choice of the heterogeneity distribution's parameters is an already known pervasive issue on duration models. A logical step further is the development of models that depart from the fully parametric paradigm. A modern account that tries to address this is the paper of Horowitz (1999). Horowitz develops a model that poses no restriction on the baseline hazards nor on the unobserved heterogeneity[19] and develops its statistical properties. Despite the importance of Horowitz's paper, there are many other ways of approaching the issue. In a forthcoming paper, Bierens and Carvalho (2007) develop a semi-nonparametric competing risks model through the use of Legender polynomials and apply it to a data set[20] regarding recidivism behavior of released prisoners in the USA.

A second interesting avenue to research further is motivated by the fact that the usual way of tying together the two latent durations and, consequently, making them dependent is too much restrictive! The problem of this approach is that than the two latent durations are necessarily positively related. Of course, this a nuisance, at least in our context. In order to overcome this issue, we can either follow van den Berg (2000) and assume that the unobserved heterogenity is no longer univariate, but a bivariate random variable consisting of two Gammas, one for each latent variable where the covariance matrix poses no restrictions on the sign of its off-diagonal elements. Another possibility is to take a different approach: incorporating treatment in the conditional distribution of the first distribution given the covariates and the duration of the second distribution. Recidivism and job search, respectively, in our present case. By way of the later approach, Bierens and Carvalho (2006) is able to confirm the existence of a treatment effect using

[19] He uses a proportional hazards setup, however.

[20] However, they use ICPSR ♯ 8875, a different data set than the one we use on the present paper.

the same data set as the one used in this paper.[21]

A final interesting avenue to follow has to do with heterogeneity of treatment's impact. Although the parameter δ serves as a general measure of program effectiveness, it is a crude measure, indeed. One of the agreements on the literature of program evaluation is that given the specificities of the groups of people who usually make use of those services, some programs that work very well for a given group could work badly for others. In other words, the effects of programs are heterogenous and this should be accounted for. From the perspective of our model an easy choice would be $\delta(X) = \exp[X'\beta]$. However, identification of the model becomes a problem! An alternative to that approach can be seen in Bierens and Carvalho (2006).

References

Abbring, J. & van den Berg, G. V. (2000). The non-parametric identification of the mixed proportional hazards competing risks model. Manuscript, Free University of Amsterdam.

Abbring, J. & van den Berg, G. V. (2003). The non-parametric identification of treatment effects in duration analysis. *Econometrica*, 71(5):1491–1517.

Abbring, J., van den Berg, G. V., & van Ours, J. (2005). The effect of unemployment insurance sanctions on the transition rate from unemployment to employment. *Economic Journal*, 115:602–630.

Beck, A. J. & Shipley, B. E. (1989). Recidivism of prisoners released in 1983. Special report, Bureau of Justice Statistics.

Bierens, H. & Carvalho, J. (2006). Job search, conditional treatment and recidivism: The employment services of ex-offenders program. Working Paper.

Bierens, H. & Carvalho, J. (2007). Semi-noparametric competing risk analysis of recidivism. *Journal of Applied Econometrics*, forthcoming.

Card, D. & Sullivan, D. (1988). Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica*, 56(3):497–530.

Devine, T. & Kiefer, N. (1991). *Empirical Labor Economics: The Search Approach*. Oxford University Press, New York.

Eberwein, C., Ham, J., & Lalonde, R. (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.

[21] Its is important to mention that we have also generalized the treatment effect by making its impact heterogenous. In fact, this is a still open question.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult ofender recidivism: What works! *Criminology*, 34(4):575–607.

Ham, J. C. & LaLonde, R. (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on trainning. *Econometrica*, 64(1):175–205.

Heckman, J., LaLonde, R., & Smith, J. A. (1999). The economics and econometrics of active labor market programs. In *Handbook of Labor Economics*, volume 3, chapter 31, pages 1865–2097. Elsevier Science, Amsterdam.

Heckman, J. & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320.

Horowitz, J. (1999). Semiparametric estimation of a proportionalHazard model with unobserved heterogeneity. *Econometrica*, 67(5):1001–1028.

Lancaster, T. (1990). *The Econometrics Analysis of Transition Data*. Cambridge University Press, Cambridge.

Lillard, L. A. (1993). Simultaneous equations for hazards. *Econometrics*, 56:189–217.

Lillard, L. A. & Panis, C. W. (1996). Marital status and mortality: The role of health. *Demography*, 33:313–327.

Maltz, M. D. (1984). *Recidivism*. Academic Press, Orlando, FL.

Meyer, B. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4):757–782.

Milkman, R. (2001). Employment services for ex-offenders, 1981-1984: Boston, Chicao and San Diego. Discussion Paper 8619, ICPSR.

Milkman, R., Timrots, A., Peyser, A., Toborg, M., Yezer, B. G. A., Carpenter, L., & Landson, N. (1985). Employment services for ex-offenders field test. Discussion Paper, The Lazar Institute.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory*. The MIT Press, Cambridge, MA, 2nd edition.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies of causal effects. *Biometrika*, 70:41–55.

Rubin, D. B. (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Education Psychology*, 66:688–701.

Schmidt, P. & Witte, A. D. (1988). *Predicting Recidivism Using Survival Models*. Springer-Verlag.

Timrots, A. (1985). An evaluation of employment services programs for ex-offenders. Master's thesis, University of Maryland, College Park.

van den Berg, G. (1999). Empirical inference with equilibrium search models of the labor market. *The Economic Journal*, pages F283–F306.

van den Berg, G. (2000). Duration models: Specification, identification and multiple durations. In *Handbook of Econometrics*, volume V. North Holland, Amsterdam.

van den Berg, G., Lindeboom, M., & Ridder, G. (1994). Attrition in longitudinal panel data and the empirical analysis of dynamic labour market behavior. *Journal of Applied Econometrics*, 9:421–435.

van den Berg, G., van den Klaauw, B., & van Ours, J. C. (2004). Punitive sanctions and the transition rate from welfare to work. *Journal of Labor Economics*, 22:211–241.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 1st edition.

Appendix A

Detailed Descriptive Statistics

Defining T_s as the job search time since release from prison, and T_c as the time of the first arrest after release from prison and noting that our observations are interval censored, see Table 1, we have the following statistics for treatments:

Table A.1
$T_c < T_s$

Variable	Observations	Min	Max	Mean	Stdev
DRUG	12	0	1	0.1666	0.3892
RACE	12	0	1	0.0833	0.2886
SEX	12	1	1	1	0
EDUC	12	0	2	0.9166	0.5149
AGE	12	19	33	23.7500	4.2022
SANDIEGO	12	0	1	0.2500	0.4522
CHICAGO	12	0	1	0.6666	0.4923
AGEFIRST	12	12	23	16.4166	3.5537

Table A.2
$T_s < T_c$

Variable	Observations	Min	Max	Mean	Stdev
DRUG	305	0	1	0.4000	0.4907
RACE	305	0	1	0.3442	0.4759
SEX	305	0	1	0.8918	0.3111
EDUC	305	0	2	1.0819	0.4477
AGE	305	18	51	27.7180	6.3390
SANDIEGO	305	0	1	0.2885	0.4538
CHICAGO	305	0	1	0.3672	0.4828
AGEFIRST	305	6	39	16.3049	4.5191

Table A3
$T_c = T_s$ (same duration interval)

Variable	Observations	Min	Max	Mean	Stdev
DRUG	52	0	1	0.4423	0.5015
RACE	52	0	1	0.3846	0.4912
SEX	52	0	1	0.8846	0.3226
EDUC	52	0	2	1.0576	0.4160
AGE	52	19	42	27.0384	5.5552
SANDIEGO	52	0	1	0.3269	0.4736
CHICAGO	52	0	1	0.3846	0.4912
AGEFIRST	52	9	42	16.1153	5.9628

Appendix B

Joint Survival Function

It follows from equations (7), (10) and (11) that:

$$
\begin{aligned}
& S(t_s, t_c|X, V, W = 1) \qquad \text{(B.1)} \\
= \; & V.\phi_s(X) \exp\left(-V \cdot \phi_c(X) t_c^{\lambda_c}\right) \\
& \times \int_{t_s}^{\infty} \exp\left[-V \cdot (\phi_c^*(X) - \phi_c(X))\, I(t_c > \tau) \left(t_c^{\lambda_c} - \tau^{\lambda_c}\right)\right] \\
& \times \exp\left[-V \cdot \left(\phi_s(X) \cdot \tau^{\lambda_s}\right)\right] \lambda_s \tau^{\lambda_s - 1} d\tau \\
= \; & I(t_c > t_s) V.\phi_s(X) \exp\left(-V \cdot \phi_c(X) t_c^{\lambda_c}\right) \\
& \times \int_{t_s}^{\infty} \exp\left[-V \cdot (\phi_c^*(X) - \phi_c(X))\, I(t_c > \tau) \left(t_c^{\lambda_c} - \tau^{\lambda_c}\right)\right] \\
& \times \exp\left[-V \cdot \left(\phi_s(X) \cdot \tau^{\lambda_s}\right)\right] \lambda_s \tau^{\lambda_s - 1} d\tau \\
& + I(t_c \leq t_s) V.\phi_s(X) \exp\left(-V \cdot \phi_c(X) t_c^{\lambda_c}\right) \\
& \times \int_{t_s}^{\infty} \exp\left[-V \cdot \left(\phi_s(X) \cdot \tau^{\lambda_s}\right)\right] \lambda_s \tau^{\lambda_s - 1} d\tau \\
= \; & I(t_c > t_s) V.\phi_s(X) \exp\left(-V \cdot \phi_c(X) t_c^{\lambda_c}\right) \\
& \times \int_{0}^{\infty} I(\tau > t_s) \exp\left[-V \cdot (\phi_c^*(X) - \phi_c(X))\, I(\tau < t_c) \left(t_c^{\lambda_c} - \tau^{\lambda_c}\right)\right] \\
& \times \exp\left[-V \cdot \left(\phi_s(X) \cdot \tau^{\lambda_s}\right)\right] \lambda_s \tau^{\lambda_s - 1} d\tau \\
& + I(t_c \leq t_s) \exp\left[-V \cdot \left(\phi_c(X) t_c^{\lambda_c} + \phi_s(X) \cdot {t_s}^{\lambda_s}\right)\right] \\
= \; & I(t_c > t_s) V.\phi_s(X) \exp\left(-V \cdot \phi_c(X) t_c^{\lambda_c}\right) \\
& \times \int_{0}^{\infty} I(t_s < \tau < t_c) \exp\left[-V \cdot (\phi_c^*(X) - \phi_c(X)) \left(t_c^{\lambda_c} - \tau^{\lambda_c}\right)\right] \\
& \times \exp\left[-V \cdot \left(\phi_s(X) \cdot \tau^{\lambda_s}\right)\right] \lambda_s \tau^{\lambda_s - 1} d\tau \\
& + I(t_c > t_s) \exp\left[-V \cdot \left(\phi_c(X) t_c^{\lambda_c} + \phi_s(X) \cdot t_c^{\lambda_s}\right)\right] \\
& + I(t_c \leq t_s) \exp\left[-V \cdot \left(\phi_c(X) t_c^{\lambda_c} + \phi_s(X) \cdot {t_s}^{\lambda_s}\right)\right]
\end{aligned}
$$

whereas we have the following for controls:

$$
\begin{aligned}
& S(t_s, t_c|X, V, W = 0) \qquad \text{(B.2)} \\
= \; & \exp\left[-V \cdot \left(\phi_c(X) \cdot t_c^{\lambda_c} + \phi_s(X) \cdot t_s^{\lambda_s}\right)\right]
\end{aligned}
$$

The integral in (B.1) can be further "simplified" as:

$$\begin{aligned}
\textstyle\int_0^\infty \quad & I(t_s^{\lambda_s} < \tau^{\lambda_s} < t_c^{\lambda_s}) \\
\times \quad & [1 + \omega(\phi_c(X)t_c^{\lambda_c} + (\phi_c^*(X) - \phi_c(X))(t_c^{\lambda_c} - (\tau^{\lambda_s})^{\lambda_c/\lambda_s}) \\
+ \quad & \phi_s(X) \cdot \tau^{\lambda_s})]^{-(\alpha+1)} d\tau^{\lambda_s} \\
= \quad & \int_0^\infty I(t_s^{\lambda_s} < u < t_c^{\lambda_s}) \\
\times \quad & [1 + \omega(\phi_c^*(X) - \phi_c(X))((t_c^{\lambda_s})^{\lambda_c/\lambda_s} - u^{\lambda_c/\lambda_s}) + \omega\phi_s(X) \cdot u]^{-(\alpha+1)} du \\
= \quad & \int_p^q [1 + \omega\phi_c(X)t_c^{\lambda_c} + \omega(\phi_c^*(X) - \phi_c(X))(q^r - u^r) + \omega\phi_s(X) \cdot u]^{-(\alpha+1)} du \\
= \quad & \frac{1}{a}[\omega\phi_s(X)]^{-(\alpha+1)} \int_p^q a[b + x + c(q^r - x^r)]^{-(a+1)} dx
\end{aligned}$$

say, where

$$\begin{aligned}
a &= \alpha \\
b &= \frac{1 + \omega\phi_c(X)t_c^{\lambda_c}}{\omega\phi_s(X)} \\
c &= \frac{\phi_c^*(X) - \phi_c(X)}{\phi_s(X)} \\
p &= t_s^{\lambda_s} \\
q &= t_c^{\lambda_s} \\
r &= \lambda_c/\lambda_s
\end{aligned}$$

Finally, note that in order for the integral

$$\int_p^q a\left[b + x + c\left(q^r - x^r\right)\right]^{-(a+1)} dx \tag{B.3}$$

to be well-defined, we must require that:

$$a > 0,\ b \geq 0,\ p \geq 0,\ q \geq p,\ r \geq 0,\ \text{and } c > -\frac{b+p}{q^r - p^r} \tag{B.4}$$