The Pennsylvania State University

The Graduate School

Department of Economics

ESSAYS ON THE MICROECONOMETRICS OF LABOR MARKETS AND CRIMINAL BEHAVIOR

A Thesis in

Economics

by

Jose Carvalho

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2002

We approve the thesis of Jose Carvalho.

Date of Signature

Herman Bierens
Professor of Economics
Thesis Adviser
Chair of Committee

Joseph Terza
Associate Professor of Economics

Nezih Guner
Assistant Professor of Economics

Joris Pinkse
Associate Professor of Economics

Susumu Imai
Special Member

Jaeyong Lee
Assistant Professor of Statistics

Robert Marshall
Professor of Economics
Head of the Department of Economics

Abstract

The subject of the present thesis is to apply microeconometric techniques to labor economics and criminal behavior. There are two independent chapters. The aim of chapter one is to build and estimate an econometric model of recidivism for the United States using data set of released prisoners from state prisons in 1983. The data set is unique in the sense that it contains information on recidivism for both federal and state crimes in contrast of past studies restricted only to state samples with few or no information about federal crimes. The paper extends the literature on the application of econometrics to criminal behavior by treating the process of recidivism as a competing risks model where dependence arises from unobserved heterogeneity. The effects of regressors are estimated in a model that considers the non-monotonic hazards feature of criminal recidivism data sets. An accelerated time model is compared to the traditional Cox proportional model. The former appears to perform better, and is shown to capture some features of recidivism behavior reasonably well.

Chapter two evaluates the effect of a reemployment program for ex-convicts on their recidivism behavior. The econometric model has to deal with the following institutional aspects of the program. First, upon release, each ex-inmate looks for a job and this duration is represented by T_s. Second, at the same release-time the ex-criminal could commit a crime and the time of recidivism is represented by T_c. Third, the difference between controls and treatments samples only occurs conditional on finding a job. Hence, the random search duration, T_s, represents the timing of receiving the treatment. Finally, the available duration data is grouped. Our paper extends the literature on econometric program evaluation by building and estimating a model that considers the conditional feature of the treatment. We use a sample from the 1985 Employment Services for Ex-Offenders Program and find that the program helps reduce criminal activity.

Table of Contents

List of Tables viii

List of Figures ix

Acknowledgments x

Chapter 1. Recidivism, Criminal Careers and Unobserved Heterogeneity 1

1.1 Introduction 1

1.2 Competing Risks Models of Recidivism 4

1.2.1 Competing Risks Models 4

1.2.1.1 The Latent Variable Approach 4

1.2.1.2 Hazard Models 11

1.2.1.3 Model Identification 15

1.2.2 Recidivism 20

1.2.2.1 The Concept and Measures of Recidivism 20

1.2.2.2 Models of Recidivism 25

1.3 The BJS Data Set on Recidivism 31

1.4 An Econometric Model of Recidivism 37

1.4.1 Economic Analysis of Criminal Behavior 37

1.4.2 The Proportional Hazards Approach 39

1.4.3 The BJS Stratified Sample 48

1.4.4 An Accelerated Time Model 52

1.4.5 A Proportional Hazards Model 68

1.4.6 Selecting a Model 73

1.5 Conclusions 74

Chapter 2. Conditional Treatment and Its Effect on Recidivism: An Econometric Model of the ESEO Program 80

2.1 Motivation 80

2.2 Econometric Evaluation of Programs 81

2.2.1 The Evaluation Problem 81

2.2.2 Social Experiments 91

2.2.3 Non-experimental Programs 98

2.2.4 Treatment Effect and Duration Analysis 107

2.3 Characterizing the Program 112

2.3.1 Antecedents 112

2.3.2 Institutional Framework 114

2.4 The ESEO Program Data Set 117

2.5 An Econometric Model of the ESEO Program 124

2.5.1 Defining a Program 124

2.5.2 Model 127

2.5.3 The Log-likelihood Function 133

2.5.4 Estimation and Inference 145

2.6 Conclusion 152

Appendix A. 157

A.1 The Model 157

A.1.1 The case $F = 0$ 158

A.1.2 The case $F = 1$ 160

A.2 Parameter Identification 161

Appendix B. 164

B.1 Joint Survival Function 164

B.2 Duration Variables 165

B.3 Probability of Drop Out for Weibull Specification 166

B.3.1 Analytical Expression 166

B.3.2 Numerical Approximation 168

B.4 Interval Probability Distributions 171

References 174

List of Tables

1.1 Heterogeneity Distribution 15

1.2 Hazard Models 19

1.3 Covariates for Recidivism 28

1.4 Descriptive Statistics 34

1.5 Parameters Estimate: Accelerated Time Model M1 77

1.6 Parameters Estimate: Accelerated Time Model M2 78

1.7 Parameters Estimate: Proportional Model M3 79

1.8 Schwarz Criteria 79

2.1 Descriptive Statistics 122

2.2 Durations for ATTRITION = 0 123

2.3 Logit Parameters Estimate: 155

2.4 Parameters Estimate: 156

B.1 Length of Search Time - T_s 166

B.2 Recidivism Time - T_c 167

B.3 Common Intervals 168

List of Figures

1.1 Criminal Justice System 22

1.2 Misdemeanor Arrests 36

1.3 Felony Arrests 36

1.4 Misdemeanor Hazard 37

1.5 Felony Hazard 37

1.6 Function $\frac{t}{e^{X'\zeta+t^2}}$ 56

2.1 Sequencing at the JTPA 109

2.2 Sequence of Services on the ESEO Program 115

Acknowledgments

I am most grateful and indebted to my thesis advisor, Dr. Herman Bierens, for the financial support, superb guidance, endless patience and continuous encouragement throughout the time we spent together. I also thanks my other committee members for their comments. I am also thankful to Dr. Neil Wallace for his guidance and financial support. The financial support from CAPES, Brazil is kindly appreciated. I am very lucky to have four women who love me: Sofia, Cecília, Lia (Maria Auxiliadora) and Georgina. I hope that in the near future Sofia and Cecília could understand why "dad has not been playing with us for the whole week !?" The time I did not spend with them will never return, but hopefully it will guarantee them a better life. I own a great debt to my mother, Georgina, she has always been by my side. I am immensely indebted to my wife, Lia. She has shown strength, love and kindness during all this four years. Without her, it simply would not be possible. Please, Lia, take this as the strangest declaration of love.

Chapter 1

Recidivism, Criminal Careers and Unobserved Heterogeneity

1.1 Introduction

In 1998, accordingly to a press release from the Department of Justice, the recidivism rate in the United States showed great variation. While the state of Montana showed the lowest rate of 11%, Utah reported the highest rate of 67%. Between those two extremes, as illustrative examples, New York appeared with 43.8%, Florida recorded 18.8% and Illinois closed that year with 39.9%. Although these figures are difficult to compare, mainly because of different criteria to measure recidivism, it is still a relevant issue, specially for those concerned with the justice system. Recidivism, as observed by Maltz (1984), can be understood as a sequence of failures: failure of the corrections system in "correcting" the ex inmate, failure of the ex inmate in being able to live in a society and failure of the society in completely reintegrating the ex inmate into a law abiding environment. Besides the important psychological, sociological and criminological impacts related to recidivism, there are also economic effects, for instance, the forgone labor market earnings, the costs of keeping inmates in prison and jails, the obsolescence of the inmate's human capital because of incarceration. It is not surprising that economists, and specifically econometricians have turned their attentions to issues related to recidivism.

From an econometric perspective, the process of recidivism is best approached by the use of survival models. That was the route followed by Schmidt and Witte (1988). They estimated a set of models of recidivism using state data from the North Carolina Department of Corrections. They concluded by noting that the proportional model performed better both in terms of fit as well as prediction than other models used by criminologists, sociologists and statistician. Also, they asserted that the tools of econometrics could and should worthily be used to issues related to criminal recidivism. Since then, it appears that the econometric approach to recidivism has not followed the development of the microeconometrics of survival analysis. From that perspective, we aim at contributing to the literature of recidivism by extending Schmidt and Witte (1988) paper into several directions.

First, we allow the possibility that the released prisoner could commit two different categories of crime: violent and non-violent. This new feature extends past models that aggregate all different types of crime into only one category. Such common assumption has been used more for convenience than for realism. Of course, issues of estimation and model identifiability play a crucial role now. We build a full parametric model and estimate it by maximum likelihood. Two models are estimated: an accelerated hazard model and the widely used proportional hazard model, also knows as the Cox model. Both models have baseline hazards that are polynomials. This is so in order to capture the time dependence of the recidivism behavior. As our results demonstrate, having two sets of crimes appears to be a sound strategy when estimating models of recidivism. Also, and more interestingly, our accelerated model performs better than the proportional model does. We discuss some possible explanations for this and hope to convince the reader that the use of accelerated models appears to capture better the essence of the process of recidivism. This simple exercise of model selection was our second contribution.

As another contribution, unobserved heterogeneity is included into the econometric model. The presence of heterogeneity, both observed and unobserved, is a inherent feature of micro data sets. So, it seems natural to try to asses its impact on the recidivism process. As far as we are concerned, our paper seems to be the first one to address this issue in the criminological literature. Even though simplicity was the key factor when deciding the best way to incorporate unobserved heterogeneity, we believe that our results demonstrates the importance of heterogeneity in this class of models. We found a very significant, however small effect for the unobserved heterogeneity.

A final distinct feature of our paper is the fact that we used a much better data set than in past studies, including Schmidt and Witte (1988). The Bureau of Justice Statistics (BJS) data set is a nationally representative data set about recidivism, and this could make a difference, specially in the robustness of econometric estimates.

The paper begins with a literature survey about competing models of recidivism in Section 1.2. There, two strands of literature are reviewd: survival models and recidivism. In Section 1.2.1, survival models whose failure could happen for more than one cause are discussed. The literature on recidivism is the focus of Section 1.2.2. We focus mainly on a particular modelling strategy of competing risk models that uses a latent variable approach to model failure time. Regarding recidivism, issues of definitions, measurability and, especially, choice of regressors receive much attention. Section 2.4 presents the Bureau of Justice data set, its strengths and limitations. Also, the chosen regressors are presented, as well as the histograms for the hazard functions and unconditional probability of failure are graphically shown for both violent and non violent crimes. The models are outlined in Section 2.5. Following the traditional route, we present the multiple proportional model as it appears in van den Berg (2000). After this, the accelerated time model is built and estimated in Subsection 1.4.4. For comparison purposes, a Cox proportional model is also estimated in Subsection 1.4.5. Finally, we conclude in Section 1.5, and offer some possible ideas for future development.

1.2 Competing Risks Models of Recidivism

1.2.1 Competing Risks Models

1.2.1.1 The Latent Variable Approach

The traditional approach to model competing risks is to use a latent variable interpretation for the multivariate distribution of failure time [1]. Earlier studies can be found in Gumbel (1960), in the two-component system of Freuden (1961), and in Moeschberger (1974). Such seminal literature originated in the areas of reliability of systems and survival time analysis in health and biology. The risk structure is multivariate with as many dimensions as the number of mutually exclusive possible types of failures. Since in many real life situations one can not observe each type of failure separately, e.g, the different possible causes of death; the lifetime of each individual component in a machine that will eventually break down; possible reasons for ending a period of unemployment[2]; each

[1] Whenever there is no confusion, we use the terms survival time and failure time interchangeably throughout the paper.

[2] Out of the labor force, retirement, work at home and work for pay.

individual survival time is interpreted as if it belongs to an unobserved random vector of survival times. The failure can happen because of any of the possible types or causes of failure. Whatever happens first, the system is said to have failed because of that specific cause. This lack of observability of the paths of each component of the random vector and the posterior knowledge of what caused the failure explains the term "latent" duration. So, one only observes the minimum time of failure and the type or cause of failure.

More formally the model consists of a vector of latent failure time $\mathbf{T} = (T_1, T_2, \ldots, T_N)$, where the index denotes the cause of failure. Hence, the survival time is just $T = \min(\mathbf{T})$. If we define $c = \arg\min_j (T_1, T_2, \ldots, T_N)$, the vector (T_c, c) gives us the observable part of our model. A common feature of the above-mentioned studies was the fact that the probabilistic content of their models came from the specification of a joint survival function, instead of the hazard function. Without loss of generality, we restrict the random vector to be two dimensional. Generalizations to higher finite dimensional vectors pose no difficulty. Let $\mathbf{t} = (t_1, t_2)$, $\mathbf{t} \in \mathbb{R}^2$. Assuming that the random vector $\mathbf{T} = (T_1, T_2)$ has a well defined distribution function, the survival function can be written as:

$$S(\mathbf{t}) = Prob(T_1 > t_1, T_2 > t_2) \tag{1.1}$$

In Gumbel (1960), the survival function is:

$$S(\mathbf{t}) = e^{-\lambda_1 t_1 - \lambda_2 t_2 - v t_1 t_2}$$

Where $\lambda_1 > 0$ and $\lambda_2 > 0$ and $v \in [0, \lambda_1 \lambda_2]$. The parameter v measures the dependence between the different survival times. In Freuden (1961), as outlined by Crowder (2001), there were two components with initially independent survival times, T_1 and T_2, exponentially distributed. However, if one component fails, the survival time of the other component, and then the whole system, is altered due to overload. The dependence comes from the fact that if a component fails, say component 1, then the parameter λ_2 of the other component's survival function changes. The survival function for $t_1 \leq t_2$ is then:

$$S(\mathbf{t}) = (\lambda_1 + \lambda_2 - \mu_1)^{-1}\{\lambda_1 e^{-(\lambda_1+\lambda_2-\mu_1)t_1 - \mu_2 t_2} + (\lambda_2 - \mu_2)e^{-(\lambda_1+\lambda_2)t_2}\}$$

It looks like that accordingly to our definition the latter model is not a latent model since the total failure of the system only happens after both components break down. However, it is possible to redefine the model into the latent framework. It is latent if the model is thought of as having two phases: in the first phase there are two possible causes of "failure", either component can break down first and trigger the second phase, which is just a univariate survival model. Another interesting example which is a special case of Freuden (1961) can be seen in Gross, Clark, and Liu (1971). They model the survival time of an individual who lost one of a pair of organs such as lungs or kidneys.

Since we can not observe the realization of all failure times in a latent model, knowledge of the survival function only is not enough to fully characterize the model. The important probability density function is the one that represents the random variable $\min(\mathbf{T})$. For $j = 1, 2$. So, given the survival function below:

$$S(\mathbf{t}) = Prob(T_1 \geq t_1, T_2 \geq t_2) \tag{1.2}$$

For $j = 1$, the desired probability density function evaluated at $t = t^*$ is just[3]:

$$S_1^{'}(t^*) = -\frac{\partial S(\mathbf{t})}{\partial t_1} |_{t_1 = t_2 = t^*} \tag{1.3}$$

If the entire sample under study was composed only of observations that failed for whatever cause, equation (1.3) is all that is needed to conduct maximum likelihood. However, such data sets would be uncommon. Indeed, survival analysis is usually subject to either truncation or censoring in some of its observations.

[3]This can be seen by noting that $P(T_1 \geq t_1, T_2 \geq t_2) = S(t_1 | T_2 \geq t_2).S(0, t_2) = [1 - P(T_1 \leq t_1 | T_2 \geq t_2)].S(0, t_2)$.

Sampling truncation[4] occurs when, for some subsets of the population, the probability of not being sampled is unknown and can not be consistently estimated. A good example is a study of recidivism where the sampling is conducted by gathering information at jails and prisons on each person arrested during a specific time interval. The population consists of every person who has ever committed a crime before. Let (Y, I, X) be a vector of outcomes of interest. The variable Y is the failure time, or time until the commission of a new crime and to get caught. The indicator I takes values 1 or 0 if recidivism occurs or not, respectively; and the vector X represents regressors. The sample formed by the set of people who does not commit a new crime, $\widetilde{\mathcal{C}} = \{(Y, I, X)|I = 0\}$, is not negligible and is not known nor can be consistently estimated. That set will never be sampled, either. So, part of the population can not be sampled and the probability of not sampling this subset of observations is unknown and can not be consistently estimated.

Censoring, instead, is different. Sampling censoring is said to exist when subsets of the population can not be sampled, but the econometrician either knows or can consistently estimate the probability of not sampling this subset of observations. For instances, a sample of ex inmates is followed over a period of time, t_{follow}, the follow up period, in a study of recidivism. After the follow up period, some people have been rearrested but others have not. In this example, recidivism time is a truncated random variable whose truncation point is t_{follow}. People with recidivism time higher than t_{follow} can not be observed. Even though we do not know when these people will be arrested or if this will ever happen, partial information is available. The probability of not being sampled can

[4] A good discussion about truncation and censoring with an emphasis on selection bias is Heckman (1999).

be easily estimated. Censoring is by far easier to find than truncation in survival data sets.

One critical point emphasized by Lancaster (1990) is that the sampling scheme has an important role to play on the distribution of the data, and, consequently, on the estimation process. Accordingly to that author, we have a **whole population** sampling, when one randomly samples the entire population regardless the cause of the failure or state[5] occupied by the unit of observation at specific time. A **flow** sampling is obtained by means of sampling a population of people either entering or leaving a state. For instance, a study of unemployment duration that follows a group of individuals fired in the same week. Finally, when one samples a set of people who occupy the same state at a particular time, there is **stock** sampling. This last form of sampling is well known to possess an intrinsic defect called length bias: those individuals with lower hazard have a higher chance to be selected, since they remain longer in the population.

One way to overcome the possible caveats that arise from sampling in survival models is to conduct the research by means of flow sampling. Also, it must be noted that the follow up period must be chosen in a way independent of the realization of each cause specific failure time. If that is the case, a full parametric approach may be feasible[6]. Given a follow up period length of t_{follow}, the likelihood can be built by using equation (1.3) for non-censored observations and equation (1.1) for the censored observations. In

[5]The term "state" is common in Economics and other Social Sciences and its meaning is analogous to the cause of failure of a machine, say. For instance, a study of duration of marriage can end up by moving to at least 2 different states, divorced or widowhood.

[6]This is one of the simplest type of censoring. Much less stringent censoring schemes can be easily handled. We postpone a deeper discussion of censoring until the issues about model estimation.

the non-censored case, let the realization of T_j be $t_j \leq t_{follow}$. Then, letting c be a dummy variable equal to zero if there is no censoring and zero otherwise, n be the sample size, and i indexes observations, one can build the following likelihood function:

$$\mathbf{L} = \prod_{i=1}^{n} \left[\left(S^{(i)}(t_{follow})\right)^{c} \cdot \left(S_j^{'(i)}(t_j)\right)^{1-c} \right] \tag{1.4}$$

More recent studies using the latent approach are Rhodes (1986) and Escarela, Francis, and Soothill (2000) in criminology, Ebrahimi (1996) and Farragi and Korn (1996) in medical research, and Katz and Meyer (1990) and van den Berg (2000) in econometrics. Despite the growing literature on the latent approach to competing risks, its development has met two majors critiques: a critique concerning the interpretation of the model and a second critique about the impossibility of identification of the model. The second critique, named the "identification crisis" by Crowder (2001) had a profound impact on the competing risks approach, and deserves a deeper development. We postpone the "identification crisis" discussion to a later subsection and discuss the interpretation critique first.

The interpretation critique is basically related to the correct interpretation of each of the cause specific failure time in $\mathbf{T} = (T_1, T_2)$. Given the model structure, marginal and conditional densities do not always find a reasonable empirical interpretation. For instance, let $f(t_1, t_2)$ be the probability density function of $\mathbf{T}$, let $f(t_1)$ be the marginal pdf with respect to T_1 and $g(t_1|t_2)$ be the corresponding conditional probability density function. The source of the critique is twofold: there are situations where the removal of a cause is impossible, such as trying to figure out the death time from cancer if the event of dying by heart attack is zero ($g(t_{cancer}|t_{heart}=0)$); another situation arises when the removal of one or more causes change completely the survival time of the reminiscent cause; and finally the removal of a cause can make the survival time of other causes meaningless, such as trying to asses the time an unemployed worker will keep on seeking jobs if the possibility of finding a job is precluded. This set up is specially important to economics and other social sciences because eliminating a possible destination will likely alter people's behavior. Another example is trying to check the meaning of making inferences based on marginal pdf's. Reconsidering the example of dying either by heart attack or cancer, let $F(t_{cancer})$ be the marginal cumulative distribution function of dying from cancer: $F(t_{cancer}) = \int_0^\infty f(t_{cancer}, t_{heart})d(t_{heart})$. Clearly the function has any practical meaning only if the integration is restricted to the set $\mathcal{C} = \{(t_{cancer}, t_{heart})|t_{cancer} \leq t_{heart}\}$. Although these critiques deserve some attention, the literature seems to reveal that a more pragmatic view towards latent variables has been advocated. A view that, accordingly to Lancaster (1990), has seen latent durations as mathematical fictions that helps to build models and not as entities of real interest.

1.2.1.2 Hazard Models

The survival function is not the only way to construct models of failure times. A different, but equivalent avenue, is to specify the model by means of its hazard function. The equivalence between survival function and hazard function is easily seen by the following:

$$\theta_j(x, t^*) = \frac{P_j^{'}(t^*)}{1 - P_j(t^*)} \tag{1.5}$$

Where $\theta_j(x, t^*)$ is the cause j hazard function, $P_j^{'}(t^*)$ is the pdf of the survival time of cause j and x is a vector of regressors. This probability density function can be obtained from equation (1.3) by integrating out T_{-j}; and $P_j(t^*)$ is the cumulative distribution function for cause j and can also be obtained from equation (1.2) by the same processes of integrating out T_{-j}. Hence, either the knowledge of the specific hazard functions or the survival function suffices to fully characterize the model. However, there are at least two advantages in using the hazard approach: the inclusion of time-dependent regressors is easier and unobserved heterogeneity (frailty, in the medical and biological literature) or random effects models can be incorporated. Now, we proceed with unobserved heterogeneity or frailty models.

The term "frailty" was first introduced by Vaupel, Manton, and Stallard (1979) in a demographic context to describe the impact of common risks on the death of people. It works as an unobserved effect or, in the setup of Panel Data estimation, as a random effect. An unobserved heterogeneity[7] term generally enters multiplicativelly into the hazard function. So, as pointed out by van den Berg (2000), a competing risks model with unobserved heterogeneity is described by:

- There are 2 latent failure time random variables, T_j for $j = 1, 2$, representing different types of causes or states[8];

[7] From now on we stick to the use of the term unobserved heterogeneity instead of frailty, since the former is the choice of the majority of econometricians.

[8] Again, extensions to any finite number of states is straightforward.

- There is a vector of time-invariant observed regressors, $X \in \mathbb{R}^k$;
- There is a time-invariant unobserved heterogeneity vector, $V \in \mathbb{R}^2$;
- Conditional on X and V the random variables T_j are independent;
- There are functions $\Theta_{0,j}$ such that for every t_j and $v_j \in V$, the hazard functions are:

$$\Theta_j(t_j|x, v) = \Theta_{0,j}(x, t) \cdot v_j, \ \forall j \tag{1.6}$$

An interesting feature of the unobserved heterogeneity model is the possibility of generating any type of dependence among the specific failure times through the specification of a suitable form of dependence on the vector V. This is so because the model is specified conditional on the unobserved vector V that must be integrated out. So, as result of the integration, there will be dependence among the failure times. Also, as made clear by van den Berg (2000), defective risks can be easily handled.

In principle, any random variable with non-negative support is a candidate to be chosen as V. However, the simplicity of the Laplace transform of V plays a major role in the choice of the final distribution[9]. The reason for that comes from the specific relationship between the survival function and the integrated hazard. It is easy to see that the survival function for cause j conditionally on regressors and specific unobserved heterogeneity is[10]:

[9]This is true for models that specify parametric V. It is fair to mention that there is a considerable number of studies that prefers a semi-parametric estimation, treating V non-parametricaly, however.

[10]See Lancaster (1990).

$$Prob(T_j > t|x, v_j) = e^{\Theta_{0,j}(x,t)\cdot v_j} \tag{1.7}$$

Since, conditionally on $X, V = (v_1, v_2)$, the survival times are independent by assumption, the conditional survival time is:

$$Prob(T_1 > t_1, T_2 > t_2|X, V) = \prod_{j=1}^{2} e^{\Theta_{0,j}(x,t_j)\cdot v_j}$$

Finally, we need to integrate out the vector V. Let $U(v_1, v_2)$ be the probability density function of V. For the sake of simplicity, the interval of integration is omitted from the equation. From equation (1.7) and the independence assumption, we obtain the following:

$$Prob(T_1 > t_1, T_2 > t_2|x) = \int\int \prod_{j=1}^{2} \left[e^{\Theta_{0,j}(x,t_j)\cdot v_j}\right] \cdot U(v_1, v_2) dv_1 dv_2$$

The last expression is just the Laplace transform of V. Some examples of commonly used V are Gamma, Positive Stable, Power Variance Function(PVF) and Lognormal. For instances, see Lancaster (1990). The Gamma distribution has been the one of most common choice for V. It has only two parameters, and a simple Laplace transform. Besides that, experience has shown that it fits relatively well the models. The family of Positive Stable distributions has also nice properties, but the Laplace transform is more complicated. The PVF family is a natural exponential family. Its main advantage lies in the fact that the Gamma and the Positive Stable distributions are special cases, thus making possible a statistical test. The Lognormal distribution does not have a closed form Laplace transform. This means that in order to integrate it out, one needs to either approximate the integral or simulate it. Nevertheless, it works well, specially in regression type models. Table 1.1 summarizes some aspects of those distributions.

Table 1.1. Heterogeneity Distribution

	Gamma	**Positive Stable**	**PVF**	**Lognormal**
Parameters	(α, τ)	α	(α, δ, θ)	(μ, σ)
Laplace Transform	$(\frac{1}{1+s\cdot\tau})^{\alpha}$	$e^{-s^{\alpha}}$	$e^{-\frac{\delta[(\theta+s)^{\alpha}-\theta^{\alpha}]}{\alpha}}$	-
Likelihood	Simple	Feasible	Feasible	No Closed Form

1.2.1.3 Model Identification

When the first papers on survival analysis using the latent variable approach began to appear, Cox (1959) drew attention to a fundamental identification problem present in such models. In a two variable set up, Cox (1959) realized that regardless of any dependence between the two variables, the model with dependent risks is observationally equivalent to a model with independent risks. This existence result was generalized by

Tsiatis (1975) for a know finite number p of variables. The Cox-Tsiatis Theorem used as a starting point the particular way that information is available to the scientist when collecting data in survival time models.

The information available to each observation in a latent model is called the **identified minimum** (j, t_j), where $j = 1, 2, \cdots N$ is the cause of failure and t_j is the failure or survival time. Of course, for censored observations we have $(0, t)$, where t is the censored time. Hence, from knowledge of only (j, t_j) it is not possible to distinguish between which of the two models is the correct one, they will both fit the data. We give a constructive proof of the Cox-Tsiatis Theorem below for $\mathbf{T} = (T_1, T_2)$.

THEOREM 1 (COX-TSIATIS). *Let $S(\mathbf{t})$ be a Latent Variable model with generic statistical dependence between failure-specific times. Then there is a **proxy model**, that is independent, $G(\mathbf{t}) = \prod_{j=1}^{N} G^{(j)}(t_j)$, and that $G(\mathbf{t}) = S(\mathbf{t})$.*

Proof

If a **proxy model** exist then it must be the case that,

$$S_j^{'}(\mathbf{t}) = \frac{-\partial G(\mathbf{t})}{\partial t_j} \tag{1.8}$$

By independence of $G(\mathbf{t})$, we obtain the following:

$$S_j'(\mathbf{t}) = \frac{-\partial G^{(j)}(t_j)}{\partial t_j} \cdot \frac{G(\mathbf{t})}{G^{(j)}(t_j)} = \frac{-\partial \log(G^{(j)}(t_j))}{\partial t_j} \cdot G(\mathbf{t}) \tag{1.9}$$

Summing over $j = 1, 2$, we obtain $s(t^*)$, the probability density function of the time of failure, regardless the cause:

$$s(t^*) = \frac{-dG(t^*)}{dt^*} \tag{1.10}$$

Integrating both sides of equation (1.10), we get $S(\mathbf{t}) = G(\mathbf{t})$. From equation (1.9), we establish the relationship between the cause specific hazard function and $S(\mathbf{t})$ and $G(\mathbf{t})$:

$$\frac{-\partial \log(G^{(j)}(t_j))}{\partial t_j} = \frac{S_j'(\mathbf{t})}{S(\mathbf{t})} = \theta_j(\mathbf{t}) \tag{1.11}$$

So, if a **proxy model** exists then by equation (1.11) one just needs to set the specific cause hazard of the independent model to $\theta_j(t)$ to obtain the desired result. Below, we show that this is indeed the case for T_1:

$$G_1^{'}(\mathbf{t}) = \frac{-\partial \log(G^{(1)}(t_1))}{\partial t_1} \cdot G^{(2)}(t_2)$$

$$G_1^{'}(\mathbf{t}) = \theta_1(t_1) \cdot e^{-\int_0^t (\sum_{j=1}^2 \theta_j(s))ds}$$

$$G_1^{'}(\mathbf{t}) = \theta_1(t_1) \cdot S(\mathbf{t}) = S_1^{'}(\mathbf{t})$$

■

Hence, any dependent competing risk model is observationally equivalent to an independent model and the identification problem appears. The literature followed two paths to cure the problem: some, as Arnold and Brocket (1983) developed restricted parametric models and proved identifiability; others like Heckman and Honore (1989) explored the potential of regressors for identification. From an econometric point of view, Heckman and Honore (1989) has more appeal since economics has a long tradition of using regressors to asses the impact of external causes on behavior.

Heckman and Honore (1989) noticed that the Cox-Tsiats result was obtained without regressors in the model. They conjectured that the use of regressors could help identify the model. They showed that under specific assumptions[11] both the proportional hazards model and the accelerated hazard model is non-parametrically identified (See Table 1.2). The key insight of the paper is a different way to look at how dependence could be generated between the cause specific survival time. Let the survival function for T_1 and T_2 be respectively $S(t_1|x) = \exp(-Z(t_1).\phi(x))$ and $S(t_2|x) = \exp(-Z(t_2).\phi(x))$. To generate random variables from an independent competing model it is necessary just

[11] Basically, continuity, differentiability and some normalizations.

to generate two independent random variables from a uniform $U(0,1)$ and then solve:

$$Z_1(t_1) = -\log(U_1).\{\phi_1(x)\}^{-1} \tag{1.12}$$

$$Z_2(t_1) = -\log(U_2).\{\phi_2(x)\}^{-1} \tag{1.13}$$

Dependence between T_1 and T_2 can be introduced by assuming that U_1 and U_2 are not independent. By assuming that the support of $\phi_1(x)$ and $\phi_2(x)$ is the whole positive reals and an additional normalizing assumption, say, that there is x_0 such that $\phi_1(x_0) = \phi_2(x_0) = 1$, identification follows.

Table 1.2. Hazard Models

Proportional Model	**Accelerated Model**
$\theta_j(x,t) = \theta_j(x) \cdot \phi_j(t)$	$\theta_j(x,t) = \theta_j(z_j(x) \cdot t)$

If the model under investigation allows one to have access to successive durations, e.g. a sequence of unemployment, then employment and finally another period of unemployment; according to Abbring and den Berg (2000b) the assumptions made by Heckman and Honore (1989) can be substantially weakened. These models are called **Successive Durations** Models. It is generally assumed that the occurrence of two consecutive realizations of the same type are independent. This is often a difficult assumption to defend. Besides that, such models are inappropriate for many situations in reliability theory or medical research. Nonetheless, such successive durations model are increasingly being used in economics and other social sciences.

1.2.2 Recidivism

1.2.2.1 The Concept and Measures of Recidivism

The word recidivism, in a criminological context, can be broadly defined as the return of a criminal behavior after an individual has been convicted of a prior offense, sentenced and corrected. It is, as emphasized by Maltz (1984), a sequence of failures; not only at the individual level but also at the many actors that form the correctional system: paroles officials, correctional programs, correctional officials and so on. From a more operational point of view, recidivism is a measure of correctional effectiveness. However, despite the apparent objectivity of the concept, the use of recidivism as an evaluation criteria is not immune to critiques.

We can divide the criticism into two lines : the ethical and moral values implicitly assumed when one uses recidivism as an evaluation tool , and operational difficulties. The first critique calls attention to the fact that there is no scientific work completely immune to personal values. Especially in the social sciences, moral and ethics play, or should play, a key role. Among the values that are common to all users of recidivism as a tool for measurement of correctional effectiveness, the following are the most common: criminals need correcting, public officials know how to correct them and, ironically, recidivism, an occurred failure, is a correctional measure. For further discussions about those issues, see Maltz (1984). So, the use of recidivism as a measure of correctional success is done by means of only the occurrence of a failure. Consequently, any positive effect of the correctional period seems to be reduced to only the rate of potential new crimes. This is at least a simplistic view of the criminal process. Although the first type of criticism is deemed important, we take a more practical stand and move on discussing the difficulties in defining and measuring recidivism.

A major source of difficulties in using recidivism as a measurement tool is due to informational constraints. For instance, regardless the definition one chooses for recidivism, say, rearrest, reconviction, or return to prison; a reasonable number of crimes are not accounted for[12]. This is due to the fact that either the criminal was not apprehended or, if apprehended, he/she was not convicted. Another example is the fact that an apparent success of a correctional program reflected by a low recidivism rate measured by rearrests, could also be a measure of success of the police force. This problem is technically an identification problem in the econometric literature, and can be overcome given available data on police force measure of effectiveness[13]. Another informational

[12]The use of self-reported data sets on criminal behavior, although in principle can overcome this type of problem, is well known to possess low reliability.

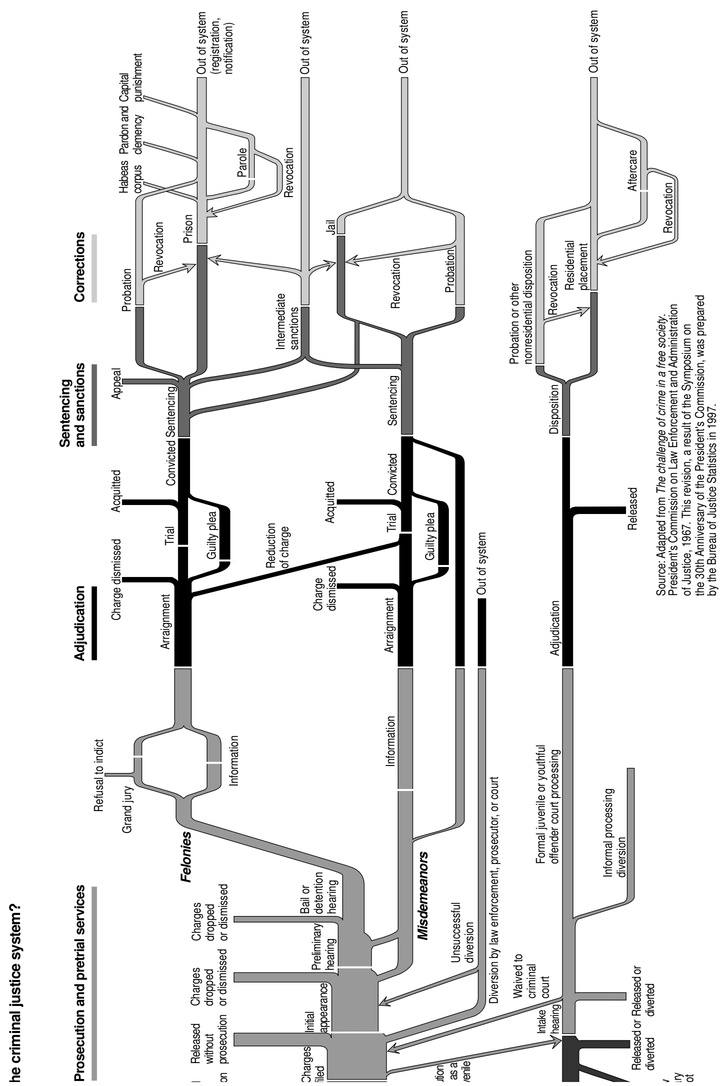
[13]It is straightforward to see that if the most common measure of police effectiveness, arrest rate, is used to disentangle the "true recidivism", endogeneity will definitely be a issue.

constraint is the fact that in some situations it is not possible to know if the event is a true recidivism or the first crime ever committed by the individual. This is the case in some states that prohibit the disclosure of juvenile information of crime behavior.

A related problem is the incompleteness of criminal justice data. This is in part a reflection of the highly complexity of the criminal justice system. This system involves all three levels of administration, local, state, and federal, and has been evolving since the creation of the country. After getting into the system by committing a crime and being caught, the individual can walk through a very elaborated process that represents different steps: law enforcement, prosecution, court , and corrections. Figure 1.1 shows a brief representation of the criminal justice system. Although not depicted, recidivism can be thought as a loop back to the beginning of the diagram. Hence, it is not a surprise that data on recidivism demands a high degree of coordination and effort to be gathered.

As there are many governmental agencies involved in the criminal justice system, the issue of availability and uniformity of data is still an issue. Specifically, correctional jail data has some room for improvement, despite great efforts from the Bureau of Justice Department. Another important issue is the lack of a data set able to incorporate both multi-state and federal criminal history. Past studies have underestimated recidivism rates because either some individuals committed crimes outside the borders of the original state or committed a federal crime. It is worth stressing that the Bureau of Justice Department has played a key role in developing a national data set on crime that considered all those issues. The data set used in our paper is much more reliable than others for reasons that will be made clear later on.

Fig. 1.1. Criminal Justice System



Any attempt to make recidivism an operational concept must pay attention to the fact that recidivism is an interval time between to events: a release event and a failure event. The release event could be from incarceration, from parole supervision, from a halfway house or any other type of official custody. So, the choice of the first event is dependent on the objectives of the study, and to a minor extent on data availability issues. The second event deserves more discussion, though. Actually, great part of the controversy in defining recidivism rests upon it. The modern tendency in criminology has shown that there are three possible definitions for recidivism: rearrest, reconviction and reincarceration[14]. 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). Also, it resembles more closely the concept of recidivism.

A sound case against the use of reincarceration or "return to prison" is that the elapsed time is not a measure of the individual criminal behavior only. It reflects also the criminal justice processing time. Only with the very heroic assumption that all crimes have the same processing time, would this measure be worth using. However, this measure may be useful for prison capacity planning or prison budgeting decisions purposes. The use of reconviction can be criticized on the same grounds as before. Hence, by elimination, we are left with rearrest as the "best" definition for recidivism. Indeed, this is the choice of scholars such as Blumstein and Cohen (1979) and Maltz (1984). Maltz (1984) concludes that *"... arrest is a better indicator of offender conduct than conviction."*. A strong appeal is the availability of arrest data in comparison to

[14]Those are just broad categories, and admit finer classifications. For a more detailed discussion, see Maltz (1984).

other types of data sets, and the fact that rearrest is an event more closely related to the individual criminal behavior. Although he made clear his position in favor of using rearrest as a measure of recidivism, he also suggested some caution for its use.

Here are the main problems in using rearrest as a measure of recidivism:

- It is possible that arrested people are innocent (Type I error);
- Arrests are much less rigorous than convictions. While *probable cause* is needed to arrest someone, *proof beyond reasonable doubt* is necessary for conviction;
- It appears that the "usual suspects " phenomenon is more than anecdotal. This leads to sampling bias, since, after a crime has occurred, police officers generally check first the department files in search for suspects to arrest;
- Sometimes police departments are put under pressure to solve crimes, usually those crimes with great impact in the media. In those cases, there is a strong incentive to make an arrest;
- When under parole, if the parolee stops reporting to the parole officer, that is, if the person absconds[15], there is no way to classify it as either a failure or not failure. Rearrests, as any other measure of recidivism, is pointless in this case. The classification of that event is totally *ad hoc*.

Despite those problems, we stick to the use of rearrests as the best available measure for recidivism. More precisely, "raw" rearrest, since there is no consideration whether or not a conviction followed the arrest.

[15]This is a technical term for "run away".

1.2.2.2 Models of Recidivism

The use of survival models to study criminal recidivism dates back to the end of the nineteen seventies. The pioneering works of Partanen (1969), Carr-Hill and Carr-Hill (1972), and Stollmack and Harris (1974) are representative of the early literature. Two decades later, the literature is able to address a host of important question through the use of survival models, such as: prediction, evaluation of programs and estimation of the effects of regressors on failure times. A good example is provided by Barton and Turnbull (1981), who evaluated a program impact on the process of recidivism, controlling for explanatory variables. Also, Schmidt and Witte (1988) is a relevant application that includes not only regressors but also a "split" parameter. It also makes clear the potential of using an econometric approach to study recidivism.

For the purposes of this paper, a model of recidivism is a survival model that specifies a relationship between the time until failure (recidivism) and the regressors that influence it. It also considers issues of censoring/truncation of the observations, potential problems of misspecification, such as unobserved heterogeneity, endogeneity, and so on. Of course, the sophistication of such models has gradually increased. An important development compared to earlier models is the explicit inclusion of regressors.

The first models that applied survival techniques to recidivism did not include regressors. This reflects in part the lack of appropriate data sets and the fact that survival models appeared first in the literature on reliability theory or medical research as a statistical model. Compared with criminology, economics, and other social sciences, those areas have paid fewer or even no attention to behavioral and/or external forces as determinants of survival models[16]. The inclusion of regressors in survival models has at least two advantages: first, the effects of observable heterogeneity can be estimated, and secondly, prediction for samples can be made conditional on observable characteristics. A good example of the importance of using regressors is given by Barton and Turnbull (1981). They show how misleading the estimation of a model without regressors can be compared to the results of the estimation of the same model with regressors. The initial difference of 100% between the rearrest rate for releasees from two Connecticut prisons vanished after controlling for observed characteristics. Chung, Schmidt, and Witte (1991) made also a clear point in favor of regressors in survival models of recidivism.

The right choice of what set of regresors is more suitable to be chosen is sometimes a process of trial and error. The determinants of what type of regressor should be included in a model of recidivism depends, of course, on what kind of beliefs one has on the relevance of each variable as a cause of recidivism. This is, needless to mention, strongly influenced by the academic background of the researchers. The interplay of criminology, sociology, psychiatry, psychology, and economics has produced a huge amount of candidates for regressors. Such amount of potential regressors can be seen in Gendreau, Little, and Goggin (1996). However, after 3 decades of using survival models to recidivism data, some common ground has been established.

The modern tendency in the use of regressors in criminology reflects the main concern about the possibilities of policy interventions in diminishing recidivism. Accordingly to Zamble and Quinsey (2001), there are two groups of regressors: static and dynamic.

[16]Of course, this reflects also the own object under study, such as the case of Reliability Theory.

The criterion of classification is basically to what extent the correctional authorities or policy makers can change the regressor. The cited authors observe that variables such as sex, age, past offences, past substance abuse are inherently non-modifiable and are examples of static variables. Other regressors, predominantly of psychological nature, such as emotions, thoughts, and perceptions, are classified as dynamics. This is so because, those regressors could be changed through public initiatives. Three aspects of the dichotomy static/dynamic from Zamble and Quinsey (2001) is worth discussing. First, there is no mention about economic regressors such as unemployment rate, wage rate, income distribution, and so on. This is not surprising, given that the rational agent approach is still seen with reserves in the criminological arena. Second, the criminological community has some concerns about variables of psychological nature mainly because the difficulty of making them operational. Finally, the dichotomy static/dynamic can be interpreted in econometric terms: the static/dynamic dichotomy is just the common exogenous/endogenous econometric dichotomy. In order to give an idea of the type of regressors used in recidivism models, we summarize, in Table 1.3, from information appearing in Gendreau, Little, and Goggin (1996) and Zamble and Quinsey (2001).

Apart from the inclusion of regressors, another important extension to models of recidivism is the inclusion of "split population" parameters. Split models, as defined by Maltz and McCleary (1977), involves the assumption that the probability of eventual failure is not necessarily one as $t \to \infty$. The probability of eventual recidivism as $t \to \infty$, for instance, would be ρ_{split}, where ρ_{split} is the splitting parameter. The new model is just slightly modified. The likelihood must account for the new conditioning that appeared. So, if one defines a dummy for the event of eventual recidivism, $R = 0$ if there

Table 1.3. Covariates for Recidivism

Static	**Dynamic**
1.Criminal History	1.Antisocial Personality
2.History of Antisocial Behavior	2.Companions
3.Family Background	3.Criminogenic Needs
4.Gender	4.Personal Distress
5.Race	5.Employment Status
6.Ethinicity	6.Future Income
7.Education	7.Interpersonal Conflicts
8.Age	
9.Income	
10.Employment History	

is no recidivism, and $R = 1$ otherwise; the probabilities that must be used in the new likelihood are: $Prob(R = 0) = 1 - \rho_{split}$ and $Prob(R = 1) = \rho_{split}$. For an observed failure the likelihood contribution must be the conditional probability of failure times ρ_{split}, and for censored observations it must be the survival probability times ρ_{split} plus $1 - \rho_{split}$. The split model can be generalized to include the effect of regressors in the splitting parameter as in Schmidt and Witte (1988). They assumed a logit model:

$$\rho_{split} = \frac{1}{1 + e^{X'\beta}}$$

There is a reasonable appeal to the use of split models in studies of recidivism. Contrary to a model of machine failure, where a component will eventually fail, or in a study of cancer survival, where the patient will eventually die, the assumption that all released prisoners will eventually commit another crime seems inappropriate. Indeed, such an assumption denies any effect of prison treatment on ex inmates. Another point in favor of split models is made clear in Schmidt and Witte (1988): different regressors affect in different ways both the splitting parameter and the time of failure, and this is important for policy purposes.

Despite the considerations made about split models, its successful use is largely dependent on the availability of data sets with very long follow up periods. Even with quite long periods in the range of $70-81$ months, as is the case in Schmidt and Witte (1988), the fit presents some difficulties. Much longer periods, such is the 21 years study used in Escarela, Francis, and Soothill (2000), seems to be statistically suitable and feasible. However, if one considers that the usual follow up periods in studies of recidivism is around 36 months, the appeal of split models is understandably diminished.

In their survey of survival analysis applied to recidivism, Chung, Schmidt, and Witte (1991) made a point about the potential of using competing risks models to recidivism. At that time, they argued that few papers had explored the approach. The few exceptions were Visher and Linster (1990) and Rhodes (1986). The former authors model the rearrest process of a sample of individuals arrested and released on their own recognizance[17]. The disposition of the case was treated as a cause specific failure[18] together with the rearrest. The latter author analyzed time until failure on parole, and treated termination of supervision without failure as a competing risk. It is worth noting that the use of competing risks model have a great potential for clarifying some important questions in criminology. For example, studies of crime escalation (increased violence in

[17]A security entered into before a court with a condition to perform some act required by law; on failure to perform that act a sum is forfeited.

[18]Note that this could be treated normally as censoring.

criminal behavior), studies of crime specialization, the very existence of criminal careers, and so on. Also, in other areas of criminology that could make use of survival analysis, the competing risks approach would be extremely useful and can easily operationalized. Those areas are tax filling time, prosecution durations, criminal career length, to name a few.

The more recent literature on competing risks model applied to recidivism seems not have received enough attention. Some example of this still growing literature are Copas and Heydari (1997) and Escarela, Francis, and Soothill (2000). As made clear by the latter authors, although of high interest for scientific and public policy purposes, *" Little attention, ... , has be given to the case where a reconviction can have several levels of seriousness or can be categorized into different types of reconviction."*. However, one can only guess the motives for the lack of studies of this type[19]. The identification problems and the complexity of models that depart from simple functional specifications have hindered the application of competing risks models to not only criminology but also to economics, sociology, medicine, operations research and so on.

1.3 The BJS Data Set on Recidivism

The data set used in our analysis comes from the Inter-University Consortium for Political and Social Research, henceforth ICPSR, study number 8875, and was originally collected by the Bureau of Justice Statistics, BJS. This data collection represents a major

[19] For the sake of completeness, it is important to mention that there have been studies using a probability transition approach to model criminal career offender on a setup of competing risks. However, those models have made little use of covariates and ignored the existence of desistence in criminal careers. A good example is Lattimore, Visher, and Linster (1994).

effort by the U.S. Department of Justice to systematically improve the measurement of recidivism. The complexity of the data gathering, its coverage and reach, makes it unique amongst other available data sources in recidivism.

After noticing that one major deficiency in information about the behavior of persons leaving prisons was the lack of nationwide data, the BJS initiated a program to overcome this problem. So, in 1983, the Bureau started a new National Corrections Reporting Program, NCRP, which follows an agent from admission to prison up to either unconditional release or successful completion of a conditional release or parole. Also, guided by another deficiency, say, the lack of data for states as well FBI data for each individual, the BJS implemented in 1985 a study to asses the viability of linking state and federal correctional data. Hence, by May, 1987, initial results were published by the BJS, and by the Winter of 1989, the ICPSR made the data set available to the public.

The BJS data set contains records of 16,355 prisoners from whom both state and FBI rap sheets were found, out of a total population of 108,580 ex inmates. The data set is a representative sample of prisoners released from 11 States[20] and who survived up to the follow-up period. Only released prisoners with sentences of at least 1 year were included. Administrative releases, prisoners who were absent without leave(AWOL), escapees, releases on appeal, transfers, and those who died in prison were excluded. So, contrary to other studies that understated recidivism, the data set contains criminal behavior information at both state and federal level. This represents an improvement in terms of national representativeness, since past studies were restricted to single states

[20]California, Florida, Illinois, Michigan, Minnesota, New Jersey, New York, North Carolina, Ohio, Oregon, and Texas.

or cities. The total number of records is 299,897, and there is information on numerous factors that affect recidivism.

As usual in social sciences, the sampling scheme was of a standard stratified type. Within each gender group in the 10 sampled States, 24 strata were created based on categories of age(less than 25, 25 to 34 and 35 or more), race(black and other races) and offense type(violent, property, drugs and other offenses), which gives a total of 10x2x24 = 480 strata. In order to account for the non i.i.d feature of the sampling, a series of weights was published in ICPSR (1989). Estimation using stratified samples without correcting for it, specially if the stratification is endogenous, is likely to produce undesirable results. A loss in the efficiency of estimators and/or inconsistency are the most common outcomes. We postpone the solution given to the stratification of the sample employed in our paper until section 1.4.4. For good references on estimation with stratified samples, see Hausman and Wise (1981) and Wooldridge (2002).

Our point of departure to choose a set of regressors of recidivism is Schmidt and Witte (1988). However, we also pay close attention to the criminologic literature in recidivism, for instance Gendreau, Little, and Goggin (1996). The regressors used in Schmidt and Witte (1988) are basically age at release, time served for the sample sentence, sex, education, marital status, race, drug use, supervision status, participation in programs, and dummies that characterize the type of recidivism. However, as pointed out by Gendreau, Little, and Goggin (1996), that set of recidivism regressors belongs exclusively to the category of static risk factors.

The modern literature on criminal recidivism distinguishes 2 clear groups of recidivism predictors: static factors and dynamic factors. Static factors are variables that belongs to the prisoner past and cannot be changed, such as age, sex, history of convictions. Dynamic factors or also known as criminogenic needs, can change and should be a target of the treatment. For example, values, behaviors , and antisocial cognitions. However, as argued by Gendreau, Little, and Goggin (1996), while the use of static factors presents no disagreement the use of dynamic factors is not a consensus in the profession, *"There is no disagreement in the criminological literature about some of the predictors of adult offender recidivism, such as age, gender, past criminal history, early family factors, and criminal associates. There has been, however, considerable controversy and/or lack of interest in dynamic risk ."*

So, given our data limitations and from what is discussed above, we choose a group of variables that belongs to the traditional static risk category. Also, and more importantly from a econometric point of view, we exclude some variables from even initial consideration because of a potential high correlation with other already included variables. This appears to be the case of variables such as past criminal history and sentence length, given the clear movement in the State's sentencing structure to tie past behavior to length of sentence. The initial variables and their justifications appear below together with a brief set of descriptive statistics:

- **TIME**: Time, in $\frac{days}{1000}$, to be rearrested or to reach the follow-up date (04/16/88) after release from prison. There is some discussion whether rearrest, reconviction or reincarceration is the best measure of recidivism. However, as pointed out by Beck

Table 1.4. Descriptive Statistics

Variable	**Observation**	**Missing**	**Min**	**Max**	**Mean**	**Stdev**
TIME	16117	238	0.0010	1.9320	0.8851	0.7024
C	16355	0	0	1	0.3449	0.4753
F	16355	0	0	1	0.2527	0.4346
SEX	16354	1	0	1	0.8677	0.3387
RACE	15972	383	0	1	0.4481	0.4973
PRISON	3076	13279	0	767	32.421	75.6531
JAIL	9442	6913	0	388	40.5964	41.696
AGE	16248	107	4.8750	26.1140	10.0673	2.9826
SENT	16194	161	0	10.9670	0.6276	0.7070
RELEASE	16218	137	0	1	0.8612	0.3457
E1	8738	7617	0	1	0.1661	0.3722
E2	8738	7617	0	1	0.7598	0.4271

and Shipley (1989), criminologists have argued that rearrest is the most reliably measure of recidivism;

- **C**: Dummy variable for the event of being rearrested, **C** = 0, or no recidivism up to the follow up end date, **C** = 1;

- **F**: Dummy variable representing the motive for the rearrest. **F** = 0 if it is a misdemeanor, not known, or others and **F** = 1 if it is a felony or felony/misdmeanor. One would argue that the dichotomy should be property crime versus non-property crimes, instead of felony versus misdemeanor, but we think that a scale of aggressiveness is more impostant;

- **SEX**: Dummy variable for sex. **SEX** = 1 for male and **SEX** = 0 for female;

- **RACE**: Dummy variable for race. **RACE** =1 for Black and **RACE** = 0 for White , Indian, Alaskan, Asian or Pacific Islander;

- **PRISON**: Prior time spent in prison by the date of admission of the last sentence before the current release, in months;

- **JAIL**: Prior time spent in jail by the date of admission of the last sentence before the current release, in months. The difference between jails and prisons is basically that offenders who commit less serious crimes and get lower sentences, usually less than 1 year, go to jail. Prisons are the final destination for serious offenders;

- **AGE**: Age in $\frac{days}{1000}$ at the beginning of the last sentence before the current release;

- **SENT**: Time served, in $\frac{days}{1000}$, in the last sentence before the current release;

- **RELEASE**: Dummy variable representing the type of release from prison. If **RELEASE** = 0, the release was under either parole or probation. **RELEASE** = 1, otherwise.

- **E1**: Dummy for educational attainment. **E1** = 1 if the highest academic grade level completed by the offender, before being admitted to prison to serve the last sentence before the current release, is Eighth grade or less, and **E1** = 0 otherwise;

- **E2**: Dummy for educational attainment. **E2** = 1 if the highest academic grade level completed by the offender, before being admitted to prison to serve the last sentence before the current release, is any High School education, and **E2** = 0 otherwise;

The data description appears in Table 1.4. To give a general idea of what kind of data we are about to use as input into our econometric model, Figures 1.2, 1.3, 1.4 and 1.5 show the histograms of arrests and unconditional hazards for both misdemeanor and felonies.

Fig. 1.2. Misdemeanor Arrests

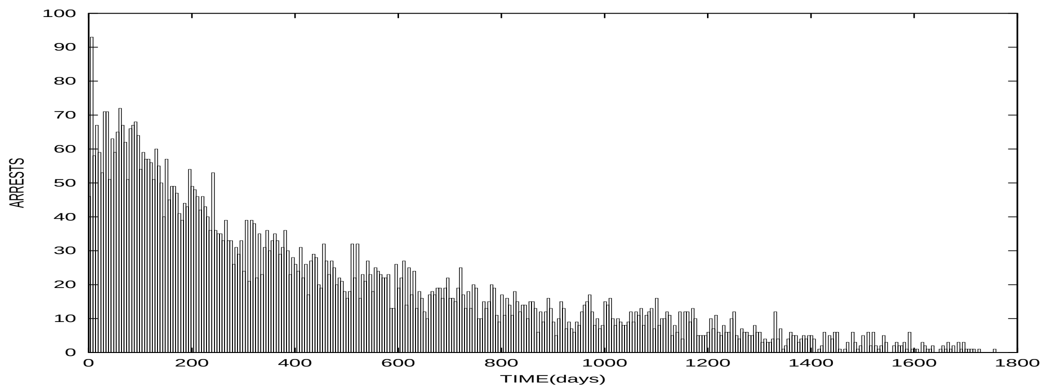


Fig. 1.3. Felony Arrests

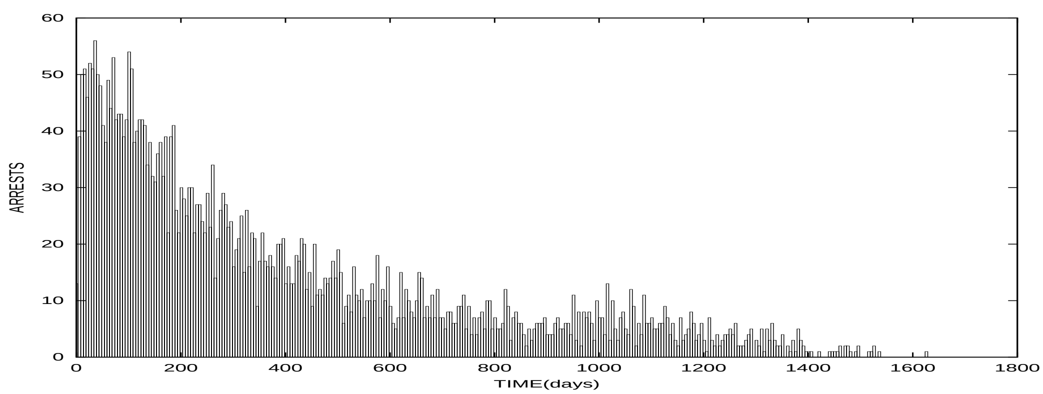


Fig. 1.4. Misdemeanor Hazard

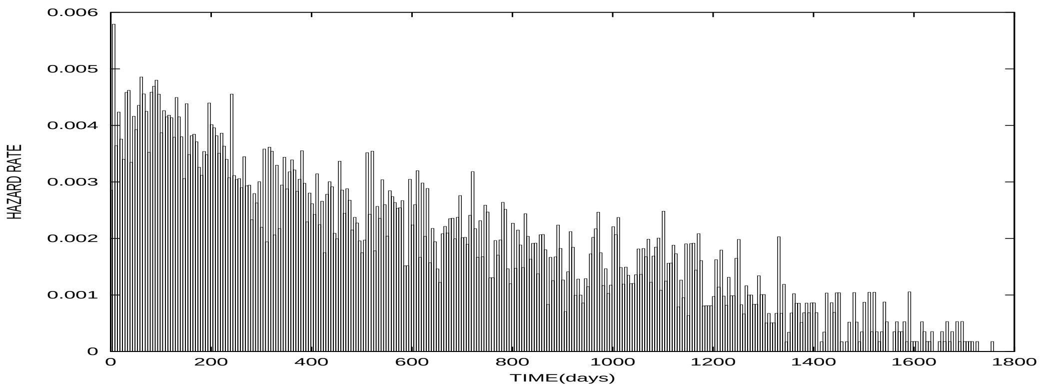
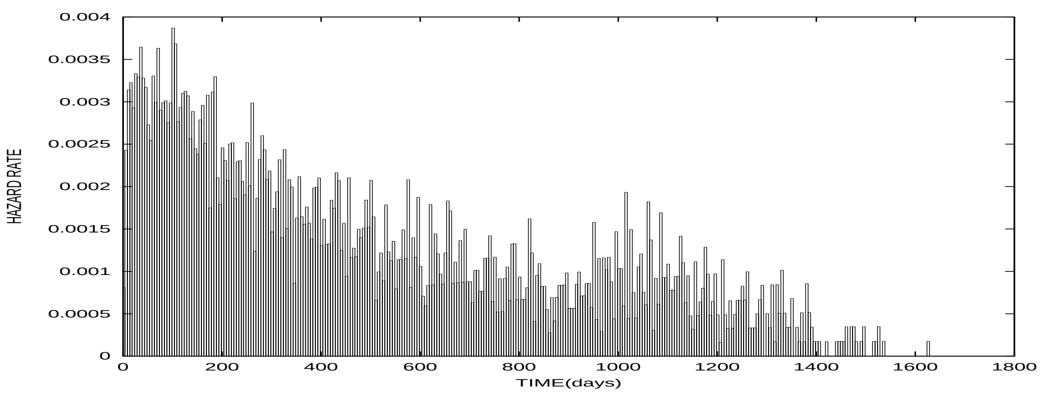


Fig. 1.5. Felony Hazard



1.4 An Econometric Model of Recidivism

1.4.1 Economic Analysis of Criminal Behavior

Since the seminal paper of Becker (1968) the economic approach to criminal behavior has incorporated the rational agent paradigm into increasingly sophisticated models. Becker's model, extended by Ehrlich (1973), advocates that the decision to commit a crime could be thought of as an additional alternative available to the agent when deciding to allocate his/her time. A very simple version of a such type of model describes a rational agent who decides every period of time to either work for pay or commit a crime. The returns of working are wages and the returns of committing a crime is a lottery between the outcome of the crime or the disutility of being in a prison[21]. The optimal choice is just the one that maximizes the expect utility. The expected value is calculated by considering the probability of being caught while committing a crime. The analysis of this type of model leads to a host of behavioral implications, some testable, others not.

From our perspective, there are two important contributions of the economic approach of criminal behavior: first, a set of regressors is justified as important to explain criminal behavior, such as age, educational attainment, wages, police effectiveness, and so on. Second, by offering a precise behavioral framework to criminal behavior, these models follow the recent trend of constructing well defined microeconomic models to explain behavior. Indeed, each cited contribution has prompted the development of two

[21]The disutility of being caught can be both phycological, or economic. Forgone earnings and future labor market stigma effects are examples of the latter.

approaches for econometric estimation of models of criminal behavior: the “reduced” form approach and the structural approach.

The “reduced” form approach uses the criminal model as a guide for the selection of important regressors and as a source of important suggestions on how to choose the best functional form of the econometric model. In other words, its use of economic theory is less explicit than in the structural approach. This approach of econometrics has its roots in the Cowles Commission research program. Earlier papers representative of this particular approach are: Myers (1983) and Witte (1980). More recent papers are: Levitt (1997) and Mocan and Rees (1999). The structural approach is explicitly based on a careful parameterization of preferences and technology of the model. Its emphasis is on the identification and posterior estimation of very detailed microeconomic models, and its ambition is, accordingly to Heckman (2000), to obtain estimated parameters that are “policy invariant” and to ascertain the effects of policy changes. A recent good example of such approach is Imai and Krishna (2001).

Although a structural approach would be feasible, our model estimation will not be of this form. There are several reasons for our choice. First, our main focus is the generalization of a already reasonable complex econometric model, so we have to trade-off between a more detailed economic model and a more elaborated statistical framework. Second, we believe that even though the economic approach of criminal behavior has its merits, there is a lot of impulsiveness involved in a large of criminal acts. Finally, we agree with Heckman (2000) in his assessment of the achievements of the structural approach to econometrics: *“The empirical track record of the structural approach is, at best, mixed. Economic data, both micro and macro, have not yielded many stable* *structural parameters."* Hence, the models developed in the next sections will be of the "reduced" form type.

1.4.2 The Proportional Hazards Approach

Survival models have been greatly simplified by the specification of separable functional forms whose estimation and inference are much easier. The proportional hazards model due to Cox (1972) is undoubtedly the most used specification in the econometric literature. Such specification facilitates the process of estimation, since Cox (1975) himself suggested a very influential approach based on partial likelihood estimation for his proportional model that avoids the specification of the separable time component[22]. Specification tests are easy to perform by the use of residuals[23]. Also, the inclusion of time-varying regressors is very simple in a proportional model. Another point in favor of proportional models is the fact that identification has been proved by Heckman and Honore (1989) both for the univariate case and the competing risks model. At first sight it seems, therefore, natural to adopt a proportional hazard model to recidivism behavior. However, this is not a clear cut. There are some strong assumptions on the proportional hazards model that seems not to fit very well with recidivism, and actually with some other models. However, we postpone for a while a deeper discussion about this issue.

---

[22] There is a drawback, however: it is assumed that the probability of having ties on two or more observations is zero. This is a very strong assumption. One can try to overcome this issue by either dropping tied observations, which is not a sound strategy, or try to model this feature. For a brief discussion, see Lancaster (1990).

[23] Those residuals are not directly related to the likelihood function, though. Models estimated by maximum likelihood do not have well defined residuals as opposed to regression analysis. Generally, a nuisance parameter is used to represent the misspecification on the hazard function. The score for this nuisance parameter is referred as the error term. Finally, residuals are defined as the error term with unknown parameters replaced by maximum likelihood estimates. For a detailed exposition, see Lancaster (1990), especially chapter 11.

Despite the initial scepticism regarding the proportional hazards specification, it makes sense, both historically and logically, to start the construction process of a model of recidivism by examining this traditional approach. We can say in advance that our model extends the few previous econometrics attempts to model recidivism, such as Schmidt and Witte (1988), in three main directions: model specification, unobserved heterogeneity, and specification analysis. In order to deal with the issues of competing risks and unobserved heterogeneity the proportional hazards models has been generalized into a model called the Multivariate Mixed Proportional Hazard Competing Risks Model(MMPHM). The discussion below follows closely van den Berg (2000).

In a MMPHM, the failure can occur for different reasons. The different failure times are assumed to be independent conditionally on a vector of regressors and a scalar[24] random variable V. Dependence amongst failure risks arises because the random variable V represents unobserved heterogeneity. Hence, conditional only on the regressors, the failure times are not necessarily independent. Within this approach, the unconditional distribution is a mixture model, whose mixing distribution is the distribution of the unobservable V, $H(v)$.

As pointed out by van den Berg (2000), a MMPHM is fully characterized by knowledge of the following:

- There are N latent failure time random variables, T_j for $j = 1, 2, \ldots, N$, representing different types of failures or destinations;
- There is a vector of time-invariant observed regressors, $X \in \mathbb{R}^k$;

[24] This could be generalized by assuming V is a random vector. A key advantage of doing that is the richness obtained with respect to the resulting covariance amongst failure times.

- There is a time-invariant unobserved regressor, V;
- Conditional on X and V the random variables T_j are independent;
- There are functions $\psi_j, \Theta_{0,j}$ such that for every j and V, the hazard functions are of a proportional hazard type:

$$\Theta_j(t_j|x,v) = \psi_j(t_j) \cdot \Theta_{0,j}(x) \cdot V \tag{1.14}$$

- The systematic hazard functions are assumed exponential:

$$\Theta_{0,j}(x) = e^{X'\beta_j} \tag{1.15}$$

For convenience, $\psi_j, \Theta_{0,j}$ and the distribution of V, satisfy all regularity conditions[25] for the standard Mixed Proportional Model stated in van den Berg (2000).

Now, the first step to estimate the model is to obtain the expression for the multivariate survival function, conditional only on the regressors, $S(\mathbf{t}, x) = Prob(T_1 > t_1, T_2 >$

[25] Assumptions 1 - 4.

$t_2|x)$. In order to do so, the conditional survival function must be integrated out using the distribution of V, $\int_0^\infty Prob(T_1 > t_1, T_2 > t_2|x, v)dH(v)$. This task is simplified by the proprieties of the Laplace transform of random variables. If $H(v)$ is a proper or defective probability distribution on $[0, \infty)$, the Laplace transform is the function defined for $\vartheta \geq 0$ by:

$$\mathcal{L}(\lambda) = \int_0^\infty e^{-\vartheta' u} dH(u) \tag{1.16}$$

Before we proceed, define the following[26]:

$$\Lambda_j(t_j) = \Theta_{0,j}(x) \cdot \int_0^{t_j} \psi_j(u)du \text{ for } j = 1, 2 \tag{1.17}$$

$$\Lambda(t_1, t_2) = \sum_{j=1}^{2} \Lambda_j(t_j) \tag{1.18}$$

$$\Lambda_{-j}(t_j) = \sum_{i \neq j} \Lambda_i(t_i) \tag{1.19}$$

[26]From this point on, without loss of generality we assume $N = 2$.

$$\Lambda(t^*) = \sum_{j=1}^{2} \Lambda_j(t^*) \tag{1.20}$$

The symbol $\Lambda_j(t_j)$ in equation (1.17) is the integrated hazard. Since the conditional durations are independent, by means of equations (1.15), (1.16) and (1.18), the conditional survival function for the random vector (T_1, T_2) is $Prob(T_1 > t_1, T_2 > t_2 | x, v) = \exp\{-v \cdot \Lambda(t_1, t_2)\}$. Given the Laplace transform of V, one obtains the following closed expression:

$$S(\mathbf{t}, x) = Prob(T_1 > t_1, T_2 > t_2 | x) = \mathcal{L}(\Lambda(t_1, t_2)) \tag{1.21}$$

Since the MMPHM is developed in a latent variable structure, and since the available data set is a single-spell with deterministic censoring, we need to realize that estimation demands the knowledge of the densities of two random variables: The failure time of risk j and the minimum time of failure. Consistently with past notation (see equation (1.3)), we want to derive the following expressions:

$$S_j^{'}(t^*, x) = -\frac{\partial S_j(t^*, x)}{\partial t^*} \tag{1.22}$$

$$S_{min}(t^*, x) = S(t^*, t^*, x) \tag{1.23}$$

We derive these two expressions in Appendix A.1. The same results could be achieved by a different path. Through the use of the properties of the Laplace transform, Lancaster (1990) shows the following results:

$$S_j^{'}(t^*, x) = -\frac{\partial \mathcal{L}(\Lambda(t^*))}{\partial t_j} \tag{1.24}$$

$$S_{min}(t^*, x) = \mathcal{L}(\Lambda(t^*)) \tag{1.25}$$

If we moreover assume that the unobserved heterogeneity V and the vector of regressors X are independent, or $V \perp X$ and secondly that V has a finite mean, $\mathbf{E}(V) < \infty$, then the model is nonparametrically identified as outlined by Abbring and den Berg (2000b). Since the model is fully parametric, we can use full-information maximum likelihood estimation. Before we proceed with the likelihood function, it is worth elaborating more about censoring.

The data used in our study is, by the definition of Lancaster (1990) and Wooldridge (2001), flow data with right censoring. A straightforward way to deal with censoring is to treat it as an additional latent variable and estimate a new model. The new model would have 3 latent variables, felony, misdemeanor and censoring. As a consequence, this new model demands only the knowledge of $S_j^{'}(t^*, x)$ for $j = 1, 2, 3$, where $j = 3$ for any observation means that censoring occurred. Treating censoring as an additional risk is suggested by Lancaster (1990) as a sensible way to deal with the problem. Yet, I think this approach introduces unnecessary complications, given the censoring of our data set. Assume, as in Wooldridge (2001), that the right censoring of our sample can be characterized by a random vector $RC = (a_i, c_i)$, for $i = 1, 2, \ldots, N$, where a_i is the sample starting date, c_i is the sample censoring date and N is the sample size. The vector RC is general enough to describe the type of sample we are going to use. Since in our data there is a common censoring date for everybody and people are release at a random date in 1983, the vector RC is (a_i, c). The simplifying assumption is that conditional on the vector of covariates X the vector of latent durations is independent of the vector RC. Given the independence assumption, the contribution to the likelihood of a non-censored observation of crime j is:

$$S_j^{'}(t^*, x|a_k, c) = S_j^{'}(t^*, x)$$

If we let c^* be the time spent in the sample, the contribution of a censored observation is:

$$S_{min}(t^*, x|a_k, c) = S_{min}(c^*, x)$$

We give the details on how to find these expressions in Appendix A.1. Define the censored indicator C, such that $C(k) = 1$ if the observation whose index is k is censored and $C(k) = 0$, otherwise. Also let the dummy $F(k) = 1$ if the crime committed was a felony and $F(k) = 0$ if it was not a felony. Let m be the independent sample size. So, the log-likelihood function $\mathbf{L}$ is[27]:

$$\mathbf{L} = \sum_{k=1}^{m} \Big\{(1 - C(k)) \cdot \Big\{(1 - F(k)) \cdot \ln \big[S_1^{'}(t^*, x)\big] + F(k) \cdot ln\big[S_2^{'}(t^*, x)\big]\Big\} + C(k) \cdot \ln \big[S_{min}(c^*, x)\big]\Big\} \quad (1.26)$$

Hence, parametric estimation can be performed and usual tests, such as Wald, Lagrange Multipliers can be conducted.

[27]For the sake of notational simplicity, we omit the index k from the densities expressions.

As realized from above, the MMPHM is flexible enough to deal with many possible situations arising in survival analysis. For instances, the problem of duration dependence, which per se is of key importance, can be handled by specifying a suitable baseline hazard function. Also the inclusion of "split parameters", such as the approach advocated by Schmidt and Witte (1988), can be more elegantly handled by a suitable choice of either a defective distribution of unobservables or a functional form of the baseline hazard. Finally, beyond finiteness of the mean and nonnegative support, one can choose a variety distribution for the unobserved heterogeneity.

Given the nice features of the proportional specification, it is not a surprise the prominent role achieved by it in the survival analysis literature. Even the accelerated model with its simplicity has not experienced a general acceptance in the criminologic and econometric circles. Notwithstanding the widespread use of the proportional model, a careful analysis of its structure is necessary in order to justify a potential choice of a different model. So, we discuss the following features that usually go unnoticed when assuming a proportional specification:

- From $\Theta_j(t_j|x,v) = \psi_j(t_j) \cdot \Theta_{0,j}(x) \cdot V$, with $\Theta_{0,j}(x) = e^{-X'\beta}$ it is easy to see that $\frac{\partial \ln[\Theta_j(t_j|x,v)]}{\partial x} = -\beta$. This means that the proportional effect on the hazard rate of a change in one regressor is the same regardless of when it is measured. This amounts to assume that, in a model with fixed regressors, a change on a specific regressor will have a ever lasting constant impact on a individual;

- The only effect of regressors is to shift either up or down the hazard function, without changing the mode location. Only the mode's height is moved. See details in Kiefer (1988);

- The relative risk, $R_{i,j} = \frac{\Theta_i(t_i|x,v)}{\Theta_j(t_j|x,v)}$, is constant through time and depends again only on the regressors. Actually, it is not possible to assess the absolute risk for a specific failure time. This could be a drawback, since as exemplified by Hougaard (2000), an insurance company might be interested also in the measurement of absolute risk of, say, dying of cancer;

- Lancaster (1990) called attention to the fact that there is no structural econometric model able to generate a hazard function of the proportional type. After 10 years, the economics profession is still far from obtaining such models, unless with very stringent assumptions. This was made clear in van den Berg (2000).

1.4.3 The BJS Stratified Sample

Before we start, it is important to discuss in more details the available sample. The BJS data set is not a random sample. From each of the 10 participating states, a separate, representative sample of male and female prisoners was drawn, the exception being Minnesota, where all prisoners were selected. Then, within each group, prisoners were grouped into 24 strata that were defined based on race, age, and type of offense[28]. Finally, an i.i.d sample was selected within each stratum. So, our sample is stratified and there are 240 strata. More precisely, we are dealing with a standard stratified (SS)

[28]The offense for which incarceration ended somewhere in 1983. After release, the prisoner belongs to the population under study.

sample, as pointed out by Wooldridge (2002). A SS sample is characterized by the following:

- The population is divided into J nonempty, mutually exclusive, and exhaustive strata, $\mathcal{W}_1, \mathcal{W}_2 \cdots \mathcal{W}_J$, where J is a finite integer;
- For each each stratum J, draw a random sample of size $N_j, \quad j = 1, 2, \cdots J$. The strata size N_j is not random;
- These samples are merged into a sample of size $N = \sum_j N_j$ is collected. The intra-stratum observations are i.i.d , each distributed as $D(\mathbf{w}|\mathrm{I}_{\mathcal{W}_j} = 1)$, where the indicator function $\mathrm{I}_{\mathcal{W}_j}$ is one if $\mathbf{w} \in \mathcal{W}_j$ and zero otherwise. However, the inter-stratum observations are not i.i.d..

It has been known that sample stratification can render estimators inconsistent and/or inefficient if necessary corrections are not made. Nevertheless, accordingly to Wooldridge (2001), in some cases, estimation can proceed without any corrections.

As a solution to the stratification problem, the most common approach is to weight the observations in order to compensate for the non i.i.d feature of the sampling scheme[29]. For SS sampling the weights are calculated by using the ratio $H_j \equiv \frac{N_j}{N}$, i.e., the sample frequencies. Wooldridge (2002) develops the asymptotics of such weighted estimators. Given the knowledge of Q_j, the probability of an observation belonging to stratum $\mathcal{W}_j$, the weight used in each observation is $\frac{Q_j}{H_j}$. The weighted estimator is both consistent and asymptotically efficient. The stratification issue can be ignored if strata are constructed

[29] Besides SS sampling, there are other schemes such as variable probability sampling (VP) and multinomial sampling (MN). However, we focus only on the case of interest, SS sampling

entirely based on exogenous variables, however. If we are modelling some feature of the distribution of the vector $\mathbf{w} = (\mathbf{y}, \mathbf{x})$ by its conditional distribution $f(\mathbf{y}|\mathbf{x})$, and stratification is done by means of the vector $\mathbf{x}$ only, estimation can proceed without using weights. Wooldridge (2002) develops this estimator and prove its consistency. Also, under some conditions, the *unweighted* estimator can be more efficient than the weighted estimator.

As discussed before, we are going to estimate our models by conditional maximum likelihood. The endogenous variable is the time until recidivism and the exogenous variables belong to the set described in Table 1.4. Since the stratification is based entirely on exogenous variables, we can ignore the stratification and proceed with the estimation. However, from inspection of Table 1.4, it is clear that we are not going to condition our log-likelihood function on the entirely set of exogenous variables used to determine each stratum. Neither information about the state nor the type of offense committed will be used during estimation. So, we want to estimate a model whose sample is stratified based on a vector of exogenous variables $\mathbf{Z}$, where $\mathbf{Z}$ contains the complement of $\mathbf{X}$ together with other variables. It remains to show that the results in Wooldridge (2001) remain valid.

Let us define the available SS sample $\mathbf{w}_{i,j,k} = (\mathbf{y}_{i,j,k}, \mathbf{x}_{i,j}, \mathbf{z}_{i,k})$, for $i = 1, 2, \cdots, N$; $j = 1, 2, \cdots, J$, and $k = 1, 2, \cdots, K$. The sample size is N. Each stratum is indexed by two numbers: j and k. So, there are $J \times K$ mutually exclusive and exhaustive strata, $\mathcal{W}_{j,k}$. The sample size of each stratum is $N_{j,k}$ and the sample frequencies are $H_{j,k} \equiv \frac{N_{j,k}}{N}$. The log-likelihood to be maximized is $L(\mathbf{y}|\mathbf{x}, \theta)$. Define the *unweighted* estimator as:

$$\widehat{\theta_u} = \arg\max_{\theta\in\Theta} \left\{ \sum_{j=1}^{J} \sum_{k=1}^{K} \left(N_{j,k}^{-1} \sum_{i=1}^{N_{j,k}} L(\mathbf{y}_{i,j,k}|\mathbf{x}_{i,j}, \mathbf{z}_{i,k}, \theta) \right) \right\} \tag{1.27}$$

In order to derive a consistent estimator of θ without using the vector Z, we assume that conditional on X, the variables Y and Z are independent:

Assumption 1. $L(\mathbf{y}_{i,j,k}|\mathbf{x}_{i,j}, \mathbf{z}_{i,k}, \theta) = L(\mathbf{y}_{i,j,k}|\mathbf{x}_{i,j}, \theta)$.

This is simply a statement about the correctness of the choice of exogenous variables, namely that the vector Z does not contain variables not included in X that are relevant. So, given Assumption 1 it is easy to see that the *unconditional unweighted* estimator is consistent and efficient:

$$\widehat{\theta_{u_z u}} = \arg\max_{\theta\in\Theta} \left\{ \sum_{j=1}^{J} \left(N_j^{-1} \sum_{i=1}^{N_j} L(\mathbf{y}_{i,j}|\mathbf{x}_{i,j}, \theta) \right) \right\} \tag{1.28}$$

The *unconditional unweighted* estimator is just the *unweighted* estimator, if we redefine the sample frequencies at each new stratum as $\overline{H_j} = \sum_{k=1}^{K} \overline{H_{j,k}}$. Hence, all desirable asymptotic properties of the *unconditional unweighted* estimator follows from Theorem 4.1 in Wooldridge (2001). There is no loss of efficiency.

1.4.4 An Accelerated Time Model

In order to build a econometric model of recidivism it is necessary to consider some characteristics of criminal behavior. This means that the researcher must take a stand regarding the way a criminal behaves. The criminological literature has not yet achieved a consensus on a "model" of criminal behavior. Different explanations for criminal behavior have been proposed by criminologists, sociologists, psychologists, biologists, and economists. From an economic perspective, the models that use the rational agent approach have been receiving attention since the seminal paper of Becker (1968). Yet, as outlined before, we do not share the belief that the rational agent paradigm is the best way to model criminal behavior. Also, we do not believe that the criminal is someone who lacks any rationality, a person who acts instinctively, without any reason or concerns about the consequences of criminal acts.

Our position is to take as a premise that the criminal possesses some rationality. In another words, the criminal respond to incentives and, to some extent, is aware of the potential impacts of criminal acts. This does not come without controversy, though. Yet, we would like to emphasize that impulsiveness is important as an explanation of criminal behavior. However, we believe that the rational approach is important for understanding criminal recidivism, even though, other perspectives will also play a key role in the process of model building.

An early attempt to apply econometric tools to study recidivism is Schmidt and Witte (1988), and throughout this section we extensively refer to that paper. The reason is twofold. First, as far as we know they were the first to apply survival analysis to recidivism. Second, a major contribution of our paper is to extend the cited paper in some directions. A simplifying assumption in their paper is that the failure time was a univariate process. There is no difference between types of crime, even though it appears that the authors had access to that kind of data. We believe that this assumption is too restrictive.

If one assumes that recidivism could happen because of different crimes or failures, there is a need to justify first the reason of why this could be the case, and also there is the problem of choosing the criteria for the classification of crimes. Assuming all crimes could be grouped together is difficult given the complexity of motivations, circumstances, and potential sanctions surrounding each specific crime. The reasons to commit a murder seems to be quite different from the reasons to cheat the IRS. However, it is also possible that some criminal trait exists, and any crime looks the same in the eyes of a criminal. From a rational choice perspective, we can interpret criminal behavior similarly to the behavior of a worker choosing different careers. So, the combination of abilities, different "technologies" required to commit each crime, and different pay-offs from doing so, can explain different criminal behavior; in particular whether the criminal becomes a specialist on a particular type of crime or not. Even though, we believe that the figure of the specialist makes much more sense from a economic point of view. The empirical evidence[30] does not give a definite answer to that question, though, as pointed out

[30] These studies, however, are not recidivism studies.

by Farrington (1997). Hence, because of our belief that the specialist is a more likely outcome and because the versatile criminal is a particular[31] case of the specialist criminal , we assume different failures in our model of recidivism.

Now, a new question arises: what is the best criterion for classifying different types of crime? This is complicated by the fact that there is no unique criterion in use, and, even if this were the case, it is sometimes tricky to choose a class for a specific crime. There are two broadly used criteria to classify crimes. The first considers the degree of violence contained in the criminal act, and crimes are either violent or non-violent. Second, the economic motivation of the crime is the focus, and crimes can be either property or personal. We believe the dichotomy violent/non violent is more adequate. Practically all societies have a clear principle that gives the protection of life a unique place in their judicial systems. This is reflected in the discontinuity of the sentencing structure in the United States if one goes from a non violent crime to a violent crime. Also, enforcement agencies are much more aggressive when trying to solve violent crimes. This argumentation becomes more convincing since our measure of recidivism to be used is rearrests. So, the first feature of our model is that there is a vector of latent failures time, (T_v, T_{nv}), representing respectively violent crime and non violent crime.

A second major step in the process of building any survival model is to specify the hazard function. The proportional hazards model has been the usual choice in econometrics. However, from what was shown in the Subsection 1.4.2, that model is based

[31] Particular in a pure econometric sense: one can start with a model with two failures time and test the independence assumption.

on restrictive assumptions. As an alternative, we choose a simple ratio of polynomials. This choice must be both easy to integrate out and must mimic the non-monotonic hazards found in recidivism studies. Accordingly to Schmidt and Witte (1988), models of criminal recidivism are characterized by a non monotonic hazard rate. The hazard rate first increases, reflecting the fact that new releasees are trying to "go straight" for a little while, and as time goes by, they return to crime. Then, after the peak, the hazard rate decrease either because the individual may become more skilled in escaping from police or because of the classical selection problem in survival models where the surviving cohort is made up of individuals with a low propensity for recidivism.

We propose as the hazard function the following rational function:

$$\Theta_j(t_j|x) = \frac{t_j}{b_j + t_j^2} \text{ for } j = nv, v$$

This is a hazard function with the potential to fit the data and is a non defective one[32]. Also, the coordinates of the point of maximum are given by $t_j^* = b_j^{\frac{1}{2}}$, and $\Theta_j(t_j^*|x) = (4b_j)^{-\frac{1}{2}}$. The assumed hazard function represents a second difference from Schmidt and Witte (1988). As usual in econometrics, we include observed heterogeneity by means of a set of time-invariant regressors[33]. We make $b_j = e^{X'\zeta_j}$, a function of regressors. So, with this hazard function it is still possible to give an easy interpretation of the effect of regressors: if a parameter corresponding to a non-negative regressor is positive and the

[32] $\Theta_j(t_j|x) > 0$, for any $t \in (0, \infty]$.

[33] See Section 2.4 for details about the regressors.

regressor involved increases, the abscissa of the point of the maximum will shift to the right and the value of the hazard at the maximum will decrease, and the whole curve will flatten. The opposite will happen if the parameter involved is negative. Figure 1.6 makes this clear. There are three plots for the hazard function: a baseline function ($\frac{t}{1+t}$), a positive effect function ($\frac{t}{e+t}$), and a negative effect function ($\frac{t}{e^{-1}+t}$). The three functions correspond to the middle, lower, and upper plots respectively. The positive effect function corresponds to a positive estimated parameter, and the effect is to shift the mode to the right and decrease the hazard function. The negative effect function does exactly the opposite. It is worth to mention that such effect will die out in the long run, as is clear from Figure 1.6. These features on the hazard function differ from the proportional model. In the proportional hazard model only the height of the hazard function is affected by a change of parameters.

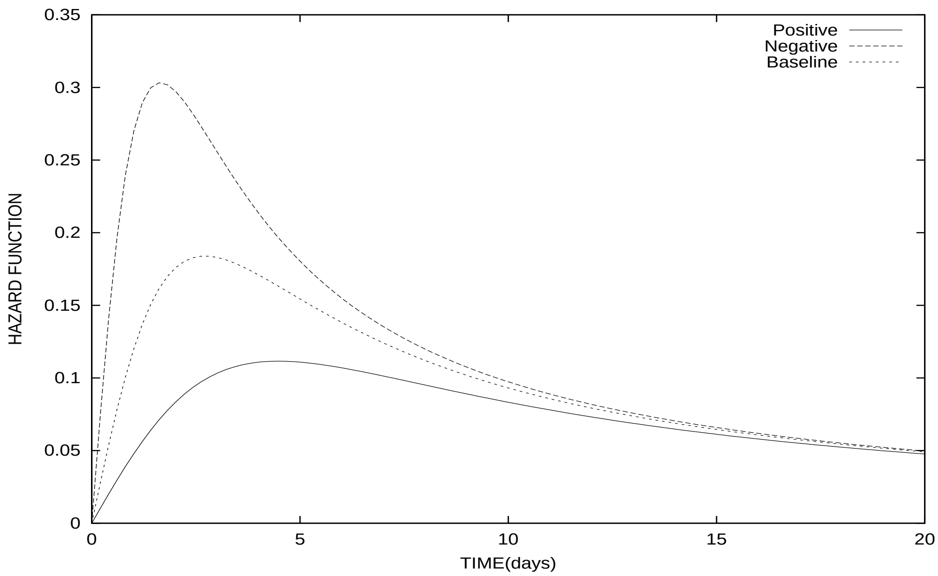
The hazard function we will use is easily seen to belong to the family of **Accelerated Time Models**. This family of models is characterized by having a hazard function of the following type:

$$\Theta(t|x) = H(t \cdot \chi(x)) \cdot \chi(x) \tag{1.29}$$

Note that:

$$\frac{t}{e^{X'\zeta} + t^2} = \left(\frac{t \cdot e^{\frac{-X'\zeta}{2}}}{1 + \{t \cdot e^{\frac{-X'\zeta}{2}}\}^2} \right) \cdot e^{\frac{-X'\zeta}{2}}$$

Fig. 1.6. Function $\frac{t}{e^{X'\zeta + t^2}}$



Moreover, note that the latent multivariate version of the accelerated model is identified, see Heckman and Honore (1989). This is a crucial feature, because as discussed before the identification of conditional multivariate survival models is not obvious.

Models of multivariate latent variables are not completely novel in studies of recidivism. A nice example with also good references is Escarela, Francis, and Soothill (2000). However, almost all those models assume that the risks are independent. We think that this is too restrictive an assumption, and, therefore we assume a specific form of dependence between risks, allowing only non-negative dependence. This is a generalization that can be tested. Following van den Berg (2000), we include unobserved heterogeneity by means of a multiplicative term into the hazard functions. This random variable V has non negative support, and finite mean, $E(V) < \infty$. The model specification becomes then:

$$\Theta_j(t_j|x, v) = \frac{t_j}{b_j + t_j^2} \cdot v \tag{1.30}$$

This is an example of dependence generated by a mixture distribution. Econometric models, and survival models are no exception, use only a limited set of regressors[34], and one always should be aware about unobserved heterogeneity. Therefore, we interpret

[34] Note that this is not due to a limitation of the model but, instead, to a limitation of the available data sets.

unobserved heterogeneity as being due to missing regressors[35]. But, this is only part of the justification. It remains to discuss possible candidates for missing variables, and the reason of having V as multiplicative term. A first candidate for V could be "external" effects, one of these external effects may be due to the structure of the sampling. All released prisoners enter the sample in the same year of 1983. This interpretation sees V as an aggregate shock that has different effects on each ex inmate. Ex inmates are homogenous in terms of failures time after considering observed regressors. Tentative examples might be a heterogeneity of police efficiency among states that persisted for a reasonable period of time, or differences in local labor markets. A second candidate for V is the traditional individual unobserved heterogeneity interpretation. There is some individual features that might have an impact on failures time whose distribution in the population is represented by V. Caution must be exercised with this interpretation, however.

We think that some variables that are difficult to measure or even to be conceptualized, such as impulsiveness, aggressiveness, or peer pressure can in principle be reasonable explanations for V. Actually, given our criterium for grouping crimes, we are sympathetic to an interpretation that goes in the direction of some psychological effect. However, we are aware of the possible misunderstandings such an interpretation could bring. Someone trying to address the effects of inherited traits on criminal behavior must be equipped of a much better data set and be prepared to step into a highly controversial issue that easily transpose the borders between science, moral and, ethics. As econometricians,

[35]Note that given our specification other traditional justifications for the existence of a mixed distribution, such as error in recorded durations, and error in recorded regressors are difficult or even impossible to develop. See Lancaster (1990), specially in Chapter 4.

we prefer to interpret V as a psychologic trait without discussing its origins. Although not necessary for our analysis, different treatment in prisons, childhood problems, and stigma effects are good candidates to give rise to that trait. Finally, the choice of a multiplicative unobserved frailty is just a matter of mathematical convenience. Without this, the Laplace transform can not be used, and the task of integrating out V will be much more difficult. The chosen distribution for $V \sim G(\alpha, \tau)$ is a Gamma distribution and its Laplace transform is given by:

$$\mathcal{L}(s) = \Big(\frac{1}{1 + s \cdot \tau}\Big)^{\alpha} \tag{1.31}$$

We are now ready to complete the construction of our econometric model. The integrated hazard (1.17) is given by:

$$\Lambda_j(t_j) = \frac{1}{2} \ln \big(1 + \frac{t_j^2}{e^{X'\zeta_j}}\big) \tag{1.32}$$

Combining equations (1.24), (1.31) and (1.32), the survival function for the failure time of risk $j = 1, 2$ conditional on being the first to happen at time t is[36] :

---

[36]The indexing subset $-j$ is defined as $\{1, 2, \cdots j-1, j+1, \cdots, n\}$. Since the present model has only two risks, $-j = \{2\}$, if $j = 1$.

$$S_j^{'}(t,X) = \frac{\alpha\tau t \cdot \left(1 + \frac{\tau}{2} \cdot \ln\left\{\left(1 + \frac{t^2}{e^{X'\zeta_1}}\right) \cdot \left(1 + \frac{t^2}{e^{X'\zeta_2}}\right)\right\}\right)^{-\alpha-1}}{(e^{X'\zeta_j} + t^2)} \tag{1.33}$$

The survival function for the minimum is obtained by combining equations (1.25), (1.31) and (1.32):

$$S_{min}(t,X) = \left(1 + \frac{\tau}{2} \cdot \ln\left\{\left(1 + \frac{t^2}{e^{X'\zeta_1}}\right) \cdot \left(1 + \frac{t^2}{e^{X'\zeta_2}}\right)\right\}\right)^{-\alpha} \tag{1.34}$$

Two models are estimated. Model $M1$ sets $\tau = 1$ and model $M2$ estimates τ together with the vector of parameters $(\zeta_1, \zeta_2, \alpha)$. This is so because the intercepts and τ can not be separately identified in the proportional model[37]. The details of the identification of τ and α are described in Appendix A.2. Also, we created the dummies $C_k = 0$, if a crime is committed, and $C_k = 1$, otherwise; and $F_k = 0$ if a committed crime is a misdemeanor, and $F_k = 1$, otherwise. Given a random sample of m individuals, and only two possible types of crimes the log-likelihood is the following:

[37]We show this result when dealing with the proportional model.

$$\mathbf{L} = \sum_{k=1}^{m} \Bigg\{ (1 - C_k) \cdot \Bigg\{ (F_k - 1) \cdot \ln(e^{X_k' \zeta_1} + t_k^2)$$

$$- F_k \cdot \ln(e^{X_k' \zeta_2} + t_k^2) + \ln(\alpha) + \ln(t_k) + \ln(\tau) \Bigg\} +$$

$$(C_k - \alpha - 1) \cdot \ln \Big(1 + \frac{\tau}{2} \cdot \Big\{ \ln \big(1 + \frac{t_k^2}{e^{X_k' \zeta_1}} \big) + \ln \big(1 + \frac{t_k^2}{e^{X_k' \zeta_2}} \big) \Big\} \Big) \Bigg\} \quad (1.35)$$

The estimation was performed using EasyReg, Easy Regression International, a free econometric software built by Professor Herman J. Bierens from PennState and Tilburg University. Among other econometric related features, EasyReg has an option for conducting maximum likelihood estimation with user supplied likelihood function. It uses the simplex method of Nelder and Mead[38] which does not rely on derivatives. We show the estimation results in Table 1.5. As shown in Table 1.5, model $M1$ performs reasonable well. Judging from the t-values, all coefficients are significantly different from zero at a 5% level[39]. The only exception is $\zeta_{(nv,sex)}$.

As demonstrated in Figure 1.6, the effect of regressors in a accelerated model is stronger in the short run than in the long run. To the best of our knowledge, this feature of accelerated model has not received great attention in criminal recidivism studies, as

[38] For details about this method and its numerical implementation see Press, Flannery, Teukolsky, and Vetterling (1989).

[39] For the sake of convenience, from now on, any reference made about parameter significance implicitly assumes a level of 5%. Also, for clarity of exposition, all results are shown with two decimals.

well as the accelerated model itself. This feature fits in the typical recidivism behavior. Accordingly to the literature in criminal recidivism, as outlined in Table 1.3, we only use static regressors. It is more likely that the impact of such regressors is stronger in the initial periods of recidivism, and then dies out after some period. Before we begin an analysis of the estimated parameters, there is a straightforward quantitative interpretation for the ζ's, the estimated parameters. Since the abscissa of the maximum of the hazard function for crime j, t_j^*, is given by $t_j^* = e^{0.5X'\xi}$, the effect of an infinitesimal[40] change in the variable X_j in the proportional change of t is just $\frac{\xi_j}{2}$. For instance, a $\xi_j = 0.7$ means that a infinitesimal change in x_j causes a 35% shift to the right in t^*.

The variables *sex* and *race* have the expected negative signs for both type of crimes. So, *ceteris paribus*, a male has a hazard function which is higher than the female's hazard function[41], at least in the short run, and the same occurs if blacks and non blacks are compared. It is important to stress that our measure of recidivism might have been playing a key role in this specific result, since the evidence of racial profiling in the USA is much more than anecdotal. The two variables do not show a similar pattern found in past studies, specially Schmidt and Witte (1988), where the effect of race is higher than the effect of sex. As expected, for violent crimes, the gender has a greater effect than the race of a person. Finally, the effect of race is remarkably similar for different types of crime.

The parameter of the variable *release* is significant with respect to both type of crimes. Although these parameters have opposite signs, $\zeta(nv, release)$ is positive and

[40] If the variable is discrete, the change is discrete, $\Delta X_j = 1$, and the interpretation is analogous.

[41] If one prefers, since hazard functions and survival functions are related, we should say that males are more likely to be rearrested earlier than females.

$\zeta(v, release)$ is negative, see Table 1.5. Consistent with other studies, the type of release has an important effect on recidivism. Since, as far as we know, there is no other attempt to fit accelerated models to multiple crime recidivism, giving a reasonable answer to that effect seems important. For non violent crimes, being completely free has the effect of deterring new crimes, whereas the contrary happens to violent crime. A tentative explanation relies on the important role played by the parole officer. This person plays an important role when deciding if actions would be counted towards a violation of either parole or probation. Also, as outlined in Maltz (1984), the parole officer exercise a strong and decisive role in controlling and following closely the parolee. Given that, an officer is more likely to tolerate non violent crimes than violent crimes. Such effect can not be identified with our available data, though. Whether such effect is a feature of our sample or of our model remains to be checked. The potential confirmation of this effect would be policy relevant information for the criminal system.

In order to analyze the effect of the variable *education*, it is worth to remember that $educ_1$ and $educ_2$ are dummies that represent maximum schooling achievements. The negative sign for both $\zeta_{(nv,educ_1)}$ and $\zeta_{(nv,educ_a)}$ are not difficult to understand. For those who have at best some years of high school, the likelihood of keeping a decent job is low. Hence, the involvement in criminal activities is a more likely event. These parameters are significant. This is in contrast to the result of Schmidt and Witte (1988), who found that formal education does not significantly affects recidivism rates. However, a reasonable explanation for why we found a positive and significant effect on violent crimes seems to be a difficult task. The age effect is quite similar to other studies. Nonetheless, the effect we found is much smaller than expected. A quick checking of

Table 1.3 will reveal that the variable *age* has one of the smallest effect across models and type of crimes. Accordingly to our interpretation of ζ, its marginal effect is not higher than 7.5%. Perhaps the harsher treatment towards juvenile crime could explain such irrelevant age effect. The controversial debate about a positive sentence length effect appears to receive some support from our results. This effect is significant and reasonable strong for both crimes. For non violent crime the marginal effect is 25.4%, and for violent crime 23.6%. Such kind of empirical evidence is likely to be important for policy purposes.

The impact of unobserved heterogeneity is highly significant, but its magnitude is relatively small. In order to make our statement more precise, the parameter ζ_{gamma} has the following interpretation. From Equation (1.30), we have that:

$$\Theta_j(t_j|x,v) = \frac{t_j}{b_j + t_j^2} \cdot v$$

Hence, $\Theta_j(t_j|x) = \int_0^\infty \frac{t_j}{b_j+t_j^2} \cdot vdH(v)$. Since $\tau = 1$, v and the vector X are independent, and $\mathrm{E}(v) = \zeta_{gamma} \cdot \tau$; a consistent estimator of $\mathrm{E}(v)$ is ζ_{gamma}. The interpretation of ζ_{gamma} follows from:

$$\Theta_j(t_j|x) = \frac{t_j}{b_j + t_j^2} \cdot \mathrm{E}(v) \tag{1.36}$$

The parameter ζ_{gamma} affects the conditional hazards by either inflating or deflating it. The abscissa of the hazard's maximum remains the same, though. From our model, $\zeta_{gamma} = 0.6789$ and again such deflating effect is stronger on the short run. Although the effect seems small, it is important to realize that if one is interested only in recidivism regardless the type of crime, such aggregate effect is far from negligible. An interesting extension would be to build a model where $v = (v_{nv}, v_v)$ is a vector of frailties. Each one multiplying each crime-specific hazard function, and with unrestricted correlation. The association measures developed by van den Berg (1997) would be an important approach for such extended model, and new insights could be gained in the study of criminal recidivism. Next, we estimate model $M2$.

Model $M2$ performs at least as good as the previous model. As shown on Table 1.6, all parameters are significant and at least the direction of the effects are the same for all variables. The only parameter not significant is $\zeta_{(nv,educ2)}$. However, three results are worth of comments. First, the parameter $\zeta_{(gamma)}$ is greater than one and is significant. Second, the parameters $\zeta_{(v,sex)}$, $\zeta_{(v,race)}$, $\zeta_{(v,sex)}$ and $\zeta_{(v,race)}$ have greater effect than the same parameters from model $M1$. Finally, the parameter ζ_{tau} is less than one and significant. Yet, the right alternative hypothesis to be tested should be $H_0 : \zeta_{tau} = 1$ against $H_1 : \zeta_{tau} \neq 1$. A simple t-test was performed and it thoroughly rejects[42] H_0. Hence, it appears that the assumption that $\zeta_{tau} = 1$ is restrictive. Also, not surprisingly the likelihood of model $M2$ is higher than the likelihood of model $M1$.

The interpretation of the unobserved effect is a little trickier, since $\mathrm{E}(v) = \mathrm{E}(\alpha \cdot \tau)$ and the parameters ζ_{gamma} and ζ_{tau} are not necessarily independent. However, at least asymptotically the unobserved effect is easy to interpret. Since the product $\alpha \cdot \tau$ is a continuous function over the set $\mathcal{D} = \{(\alpha, \tau) : \alpha \in (0, \infty), \tau \in (0, \infty)\}$, $\zeta_{gamma} \rightarrow \alpha$ and $\zeta_{tau} \rightarrow \tau$ in probability; we have the following:

$$\zeta_{gamma} \cdot \zeta_{tau} \rightarrow \alpha \cdot \tau \text{ in probability}$$

A simple application of the delta method as it appears in Lehmann (1999) gives an asymptotic test for the hypothesis that $\zeta_{gamma} \cdot \zeta_{tau} = 1$. Since the estimated covariance matrix of $(\zeta_{gamma}, \zeta_{tau})$ is $\begin{pmatrix} 0.1072 & -0.0125 \\ -0.0125 & 0.0014 \end{pmatrix}$, the asymptotically normal

[42] The value of the t-statistic is $\frac{0.15659-1}{0.03844} = -21.94094$.

statistic is:

$$\frac{\zeta_{gamma} \cdot \zeta_{tau} - 1}{(\zeta_{gamma}^2 \cdot \sigma_{tt} + \zeta_{tau}^2 \cdot \sigma_{gg} + 2\zeta_{tau} \cdot \zeta_{tau} \cdot \sigma_{tg})^{\frac{1}{2}}} = -54.3077$$

Where $\sigma_{tt} = var(\zeta_{tau})$, $\sigma_{gg} = var(\zeta_{gamma})$ and $\sigma_{tg} = cov(\zeta_{tau}, \zeta_{gamma})$. The test rejects $H_0 : \zeta_{gamma} \cdot \zeta_{tau} = 1$. Hence, there is a highly significant deflating effect of 0.29 which is higher than the deflating effect of model $M1$. For comparison purposes, the next section builds and estimates a model of recidivism using the proportional hazards functional specification. Such specification is by far the most common used not only in studies of recidivism but also in survival analysis in general. There are some similarities with our results so far, but there are also some very important differences.

1.4.5 A Proportional Hazards Model

In order to compare the relative performance of the accelerated time model, we also specify and estimate a proportional hazards model. In view of Subsection 1.4.2, we make the following assumption regarding the baseline hazards:

$$\psi_j(t_j) = \frac{t_j}{b_j + t_j^2} \tag{1.37}$$

Now b_j is a parameter to be estimated rather than a function of regressors. Note that the point of maximum is attained at $t^* = b_j^{\frac{1}{2}}$. This specification for the baseline hazards is restrictive, though. A major advantage of the proportional model is that estimation can be performed with a baseline hazard left completely unspecified[43]. Of course, the estimated effect of regressors will be partially informative, given the structure of the proportional specification. However, in order to compare the two specifications, proportional *versus* accelerated, it is desirable to have "similar" hazards.

Combining Equations (1.37) and (1.15), we have the following hazard functions:

$$\Theta_j(t_j|x,v) = \frac{t_j}{b_j + t_j^2} \cdot e^{X'\xi_j} \cdot v, \; j = nv, v \tag{1.38}$$

[43]Of course some restrictions apply, specially the one that assumes that all survival time observations can be ranked without ties. This partial likelihood approach works for both right censored and uncensored samples. For more details, see Lancaster (1990), specially Chapter 4.

The integrated hazard is given by:

$$\Lambda_j(t_j) = \frac{e^{X'\xi_j}}{2} \ln\left(1 + \frac{t_j^2}{b_j}\right) \tag{1.39}$$

Combining equations (1.24), (1.31) and (1.39), the survival function for the failure time of risk j conditional on being the first to happen at time t is:

$$S_j^{'}(t, X) =$$
$$\frac{\alpha\tau t e^{X'\xi_j} \cdot \left(1 + \frac{\tau}{2} \cdot \left\{e^{X'\xi_1} \cdot \ln\left(1 + \frac{t^2}{b_1}\right) + e^{X'\xi_2} \cdot \ln\left(1 + \frac{t^2}{b_2}\right)\right\}\right)^{-\alpha-1}}{(b_j + t^2)} \tag{1.40}$$

The survival function for the minimum is obtained by combining equations (1.25), (1.31) and (1.39):

$$S_{min}(t, X) = \left(1 + \frac{\tau}{2} \cdot \left\{e^{X'\xi_1} \cdot \ln\left(1 + \frac{t^2}{b_1}\right) + e^{X'\xi_2} \cdot \ln\left(1 + \frac{t^2}{b_2}\right)\right\}\right)^{-\alpha} \tag{1.41}$$

As before, we use the dummies $C_k = 0$ if a crime is committed, and $C_k = 1$ otherwise; and $F_k = 0$ if a committed crime is a misdemeanor, and $F_k = 1$ otherwise. Given a random sample of m individuals, and only two possible type of crimes the log-likelihood is the following:

$$\mathbf{L} = \sum_{k=1}^{m} \Bigg\{ (1 - C_k) \cdot \Bigg\{ (1 - F_k) \cdot \big[X_k' \xi_1 - \ln(b_1 + t_k^2) \big]$$

$$+ F_k \cdot \big[X_k' \xi_2 - \ln(b_2 + t_k^2) \big] + \ln(\alpha) + ln(t_k) + \ln(\tau) \Bigg\} +$$

$$(C_k - \alpha - 1) \cdot \ln \Bigg(1 + \frac{\tau}{2} \cdot \Big\{ e^{X_k' \xi_1} \cdot \ln \big(1 + \frac{t_k^2}{b_1} \big) + e^{X_k' \xi_2} \cdot \ln \big(1 + \frac{t_k^2}{b_2} \big) \Big\} \Bigg) \Bigg\} \quad (1.42)$$

Although the likelihood of the proportional model is similar to the likelihood of the accelerated model, the former can not be estimated without imposing some restrictions on either τ or one of the intercepts. The reason is simple. If we make $\ln(\tau) = \beta$, the likelihood becomes:

$$\mathbf{L} = \sum_{k=1}^{m} \Bigg\{ (1 - C_k) \cdot \Bigg\{ (1 - F_k) \cdot \big[X_k' \eta_1 - \ln(b_1 + t_k^2) \big]$$

$$+ F_k \cdot \big[X_k' \eta_2 - \ln(b_2 + t_k^2) \big] + \ln(\alpha) + ln(t_k) \Bigg\} +$$

$$(C_k - \alpha - 1) \cdot \ln \Bigg(1 + \frac{1}{2} \cdot \Big\{ e^{X_k' \eta_1} \cdot \ln \big(1 + \frac{t_k^2}{b_1} \big) + e^{X_k' \eta_2} \cdot \ln \big(1 + \frac{t_k^2}{b_2} \big) \Big\} \Bigg) \Bigg\} \quad (1.43)$$

Where $\eta_j = (\xi_{(j,sex)}, \xi_{(j,sex)} \cdots, \xi_{(j,intercept)} + \beta)$, for $j = 1, 2$. The parameters β, $\xi_{(nv,intercept)}$ and $\xi_{(v,intercept)}$ can not be simultaneously identified. For the sake of simplicity, we set $\tau = 1$, and call this model $M3$.

From Table 1.7, it is easy to see that sex and race have similar effects, in comparison to the accelerated models, for both type of crimes. Being a male has the effect[44] of multiplying the whole baseline hazard by a positive number greater than 1, and then making recidivism more likely to happen earlier. The parameter $\xi_{(nv,sex)}$ is not significant, and $\xi_{(v,sex)}$ rejects H_0 just slightly. Being black has that same effect. Their estimated parameters are also significant. The values of the parameters are similar across types of crimes. Whether this is a feature of the proportional model or not is not clear and awaits further research. The type of release has the effect of deflating the hazard function if released was under either probation and parole. This is true for both non violent crimes and violent crimes. It is interesting to note that Schmidt and Witte (1988) found a very weak and not significant effect of being released under supervision in their proportional hazards specification. The latter is closer to the results obtained in the model $M1$, since there we find a small effect for non violent crimes. However, for violent crimes, the effect is the opposite.

The variables representing educational achievement have the same effect for non-violent crimes in both $M1$ and $M2$: having at best a 8^{th} grade achievement has a

[44]Note that by the way of the specification of a proportional model, the effect of regressors work through $e^{X'\xi}$. So, a positive parameter means that the higher the value of the regressors, the grater the positive number multiplying the hazard, if the variable X is positive.

positive impact on recidivism, and having some high school years has a negative effect on recidivism for both types of crime. However, for violent crimes model $M3$ has a negative impact on crime, whereas the opposite occurs in model $M1$. The variable *age* has a small [45] positive impact on recidivism for both types of crime. The variable *sentence* has a positive effect. It is worth mentioning that both $\xi_{(nv,mode)}$ and $\xi_{(v,mode)}$ are significant. The greatest wave of crimes will happen around 730 and 493 days, respectively for non-violent and violent crimes. Finally, the unobserved heterogeneity effect is highly significant and works by deflating the hazard function, though, small. It is about one third of the effect found in the accelerated model.

The accelerated model and the proportional model present in general a good fit. However, the likelihood is higher for the proportional model than for the accelerated model. So, a more careful analysis is needed to decide between the two models. First, since criminology, as pointed out by Farrington and Tarling (1985), lacks a fully developed theory of criminal behavior, any choice between competing econometrics models becomes more difficult. Second, in our specific case, even if there is some consensus about the "best" way to model recidivism, our two competing econometrics models are non nested. The presence of non nested models poses some difficulties in testing of hypotheses. For good references, see, for example, Gourieroux and Monfort (1994), and Pesaran and Weeks (1999). The traditional approach of hypothesis testing is not appropriate for that framework. Next, we briefly justify our preference for the accelerated model as a better explanation for recidivism.

[45] Even if you take into account that our time variables were measured with units of $\frac{days}{1{,}000}$. For instance, it is much lower than the results found in Schmidt and Witte (1988)

1.4.6 Selecting a Model

Non nested hypotheses testing of microeconometric models is important because the same set of regressors is usually used to explain models whose agent's decisions are outcomes of different functional forms. This is the case, as outlined by Pesaran and Weeks (1999), of the probit *versus* logit specification, in the analysis of discrete choice; or exponential *versus* Weibull, in the survival literature. Such tests[46] have their origin in the seminal papers of Cox (1961), and Cox (1962), and have been evolving since then. However, as Heckman and Walker (1987) asserts, tests based on Cox's original work are computationally demanding and rarely used in practice. Besides that, available non nested tests have not achieved the generality of traditional nested models, such as the "Holy Trinity" (Wald, Lagrange Multiplier and Likelihood Ratio); requiring, very often, a good dose of *ad hoc* solutions.

We advocate a very pragmatic solution to the problem of model selection. First, if two models have the same set of conditioning variables X then their log-likelihood are comparable[47]. Second, following Heckman and Walker (1987), we compare the statistic proposed by Schwarz (1978) that penalizes models with many parameters. The statistic

[46] We have been using the terms "model selection" and "hypotheses testing" somewhat interchangeably. Even though they appear to be the same, there are very important differences. If, for instance, our goal is to test the proportional specification *versus* the accelerated time specification, we could end up rejecting both hypotheses. However, if we want to select the "best" model, then we will get at least one model. Hence, for our objectives it seems more appropriate to select a model.

[47] Note that a straightforward way to select a model is to estimate a new model formed by the linear combination of the two original models. Each model is multiplied by the weights ϕ and $1-\phi$, with $\phi \in [0,1]$, an additional parameter to be estimated jointly. A simple test of the null $\widehat{\phi} = 0$ suffices. However, this test has its own pitfalls. See, for instance, Pesaran and Weeks (1999).

is very easy to compute. Let L_j be the log-likelihood of model j, n the number of observations, and m_j the number of parameters of model j:

$$SC = L_j - \frac{n \cdot m_j}{2} \tag{1.44}$$

From the information on Tables 1.5 and 1.7, Table 1.8 shows that the accelerated $M1$ model is the best one.

Hence, both accelerated models appear superior to the proportional model which has been used in recidivism studies. Since multiple risk models of recidivism appears to be a quite new area of investigation, our results indicate that the interplay of different types of crime and unobserved heterogeneity in accelerated models can be a successful departure from the traditional approach in recidivism studies.

1.5 Conclusions

We think that the accelerated time model presents a modest, but important step toward the development of better econometric models of recidivism. Since Schmidt and Witte (1988), the econometric approach to this important issue has mainly focused on proportional models without distinction between types of crime. We conclude by discussing the main achievements of the current study, together, of course, with the many possible avenues for future research.

First, our results are novel with respect to the data set on recidivism from the Bureau of Justice Department, which to the best of our knowledge has not been used in this way. Moreover, previous research in this area use local data sets. Indeed, the very reason for the Bureau collecting this data is the common need in the profession for better data on recidivism, see Maltz (1984). However, there are some important limitations in the Bureau data set. From a econometric perspective, information on earnings just before incarceration and during the follow up period may be helpful. The same applies to more detailed data on educational achievement.

The second novelty is the disaggregation of crime in two types, violent and non-violent. As far as we know, our paper is the first to consider recidivism for different types of crimes. The still unresolved debate that asserts that criminals are either specialists or non specialists, is not our focus here. Although such debate is important in the long run, our point is just to recognize that violent and non violent crimes are very different types of crimes, and that this should be reflected in the model specification. One important area for future research is to use different criteria for distinguishing crimes, such as property/personal, or specifying more than two categories of crime. The latter may be difficult since some "interesting" crime categories have very low frequency in the sample, such as murder.

Another contribution to the literature on recidivism is the incorporation of unobserved heterogeneity. In fact, we used the traditional approach of specifying the hazard functions conditional on both regressors and unobserved heterogeneity, as outlined in Lancaster (1990) and van den Berg (2000). A future improvement could be to include a multiplicative unobserved heterogeneity for each type of crime with dependence between both, as suggested by van den Berg (2000) this approach will make the model more realistic, although computationally more demanding.

The final point is the model selection. The accelerated time model seems to perform better than the proportional model, and this looks like going against the common practice in survival analysis. Since Cox (1972), the proportional model has received a wide acceptance, mainly because it is easy to estimate[48]. However, the Schwarz statistic is a rather crude criteria to model selection. We view non nested hypotheses testing as a promising development for deciding between different specification of recidivism models.

The econometric approach to recidivism is a field of research yet to be explored. Whereas some seminal econometric approaches date back to the seventies, it appears that there are still many important issues to address. Economists, and more specifically econometricians, together with criminologists, sociologists, psychologists and statistician can play an important role in the future advancement of the study of criminal behavior.

[48] Actually, the search for a specification for survival models with regressors that could be estimated easily and with general baseline hazard was the chief objective of Cox.

Table 1.5. Parameters Estimate: Accelerated Time Model M1

Crime	**Parameters**	**ML estimate**	**t-values**
non-violent	sex	-0.04	-0.41
	race	-0.52	-7.31
	release	1.00	10.99
	educ1	-1.33	-9.00
	educ2	-1.03	-8.16
	age	0.09	7.53
	sentence	0.50	11.09
	intercept	-3.95	-17.27
violent	sex	-0.83	-7.05
	race	-0.55	-7.35
	release	-0.55	-5.15
	educ1	0.95	5.81
	educ2	0.50	3.57
	age	0.15	11.26
	sentence	0.47	8.73
	intercept	-3.20	-12.39
heterogeneity	α	0.67	61.50
	Log-Likelihood		-10359.30
	n		8527

Table 1.6. Parameters Estimate: Accelerated Time Model M2

Crime	**Parameters**	**ML estimate**	**t-values**
non-violent	sex	-0.42	-3.45
	race	-0.62	-7.22
	release	1.40	12.32
	educ1	-1.07	-6.14
	educ2	-0.89	-5.98
	age	0.05	3.24
	sentence	0.39	7.01
	intercept	-4.68	-16.53
violent	sex	-1.16	-8.83
	race	-0.74	-8.52
	release	-0.67	-5.47
	educ1	0.50	2.71
	educ2	0.13	0.85
	age	0.15	9.73
	sentence	0.48	7.64
	intercept	-3.34	-11.22
heterogeneity	α	1.85	5.65
	τ	0.15	4.07
	Log-Likelihood		-10243.36
	n		8527

Table 1.7. Parameters Estimate: Proportional Model M3

Crime	**Parameters**	**ML estimate**	**t-values**
non-violent	sex	0.02	0.20
	race	0.83	12.96
	release	-1.09	-13.05
	educ1	1.54	11.81
	educ2	1.18	10.65
	age	-0.08	-7.89
	sentence	-0.49	-11.72
	intercept	4.39	13.39
	mode	0.73	4.83
violent	sex	0.21	1.96
	race	0.97	14.03
	release	-0.24	-2.36
	educ1	0.69	4.86
	educ2	0.71	6.10
	age	-0.10	-8.08
	sentence	-0.43	-8.79
	intercept	3.21	9.53
	mode	0.49	5.08
heterogeneity	α	0.25	4.83
	Log-Likelihood		-10135.56
	n		8527

Table 1.8. Schwarz Criteria

Model	**SC**
$Accelerated M1$	-82838.80
$Accelerated M2$	-86986.36
$Proportional M3$	-91142.06

Chapter 2

Conditional Treatment and Its Effect on Recidivism: An Econometric Model of the ESEO Program

2.1 Motivation

During the period of $1980 - 1985$, the National Institute of Justice (NIJ) sponsored a controlled experiment to evaluate the impact of reemployment programs for recent released prisoners. Three well established programs were chosen, COERS in Boston, JOVE in San Diego and Safer Foundation in Chicago, to participate in the Employment Services for Ex-Offenders Program, henceforth ESEO. A total of $2,045$ prisoners who voluntarily accepted to participate were randomly assigned to either an experimental group or a control group. Those in the first group received, besides the normal services(orientation, screening, evaluation, support services, job development seminar, and job search coaching), special services which consisted of an assignment to a follow-up specialist who provided support during the job search and the 180 days following job placement. The control group received only normal services. The inclusion of special services was a major response to the increasing belief that some past employment programs had failed because ex inmates lost contact with their original programs.

Accordingly to Milkman (2001), the evaluation of these program services was not completely satisfactory. While the overall impact of the program in the recidivism rate is positive, the contribution of the special services appears to be negligible. However, we think that some issues related to potential selection bias were not carefully considered during the experimental design and this fact poses some doubt about the final results. The selection bias problem is subtle, but can have great impact in the final estimation results. Apparently, the initial randomization would preclude the selectivity problem. Nevertheless, this need not to be so, and only recently this issue received wider attention. Ham and LaLonde (1996) shows that if one of the outcomes of a program is a duration time, the initial randomization might not be sufficient to correctly evaluate the program's effect. More important, the first evaluation of the ESEO program did not account for the the conditional feature of the treatment. The timing of the treatment was completely neglected and, as will show later, this is a very important characteristic of the program under evaluation

2.2 Econometric Evaluation of Programs

2.2.1 The Evaluation Problem

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 nonexperimental setups. The evaluation problem has found great applicability within the economics profession and, indeed, is 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, Lalonde, and Smith (1999), Wooldridge (2002) and the extensive literature cited there. 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, specially the causal links between treatment and outcomes. In the evaluation of program's parlance, this step is known as establishing 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[1] 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 could 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, the effect of subsidies on small firms surviving and so on. Not surprisingly, generality almost always trades off with complexity and the evaluation problem has presented researchers 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 the selection problem are distinctive.

The usual assumption about the sampling scheme is the i.i.d. paradigm. This is a convenient assumption, since it rules out possible general equilibrium effects, as in Heckman, Lochner, and Taber (1998). For instance, in a large program that offers training to unemployed people, the local wage level could decrease as a consequence of a wave of post-trained workers flocking into the labor market. Those who have chosen

[1]Note that y_j can be discrete, continuous or mixed discrete and continuous.

not to participate suffer the impact of this negative externality in their wages. Actually, the i.i.d. assumption is stronger than the Stable Unit Treatment Value (SUTVA) assumption, generally used in the literature. The latter assumes only that treatment of unit i affects only the outcome of unit i. So, with this in mind and given a sample $\{(y_0^i, y_1^i, w^i)\}^i : i = 1, 2, \cdots N$, we proceed by discussing the outcomes of interest.

The two measures[2] of the effect of treatment most used in the literature are the average treatment effect (ATE) and average treatment effect on the treated (ATE_1). They are represented by the following expectations:

$$ATE = \mathbf{E}(y_1 - y_0) \tag{2.1}$$

$$ATE_1 = \mathbf{E}(y_1 - y_0 | w = 1) \tag{2.2}$$

The (ATE) measures the expected effect of treatment on a randomly person picked from the population under investigation[3]. The second measure of interest, (ATE_1), has the following interpretation: it measures the expected effect of those who actually participated in the program. Equivalently, it measures the expected difference between the outcome of a individual who participated in the program and his/her outcome in the

[2]It is worth mentioning another measure, the Local Average Treatment Effect (LATE), in Imbens and Angrist (1994) which posses good statistical properties in a instrumental variable context. See, also Wooldridge (2002).

[3]We dropped the superscript "i" from y_0 and y_1 for easy of exposition.

eventual case of no participation. A fundamental problem that arises, indeed this is **the evaluation problem**, is how to calculate either ATE or ATE_1 with observations on $y_1|w = 1$ and $y_0|w = 0$ only, or, y_1 for the treated and y_0 for non-treated. This means that observations on $y_1|w = 0$ and $y_0|w = 1$ are missing and **the evaluation problem** is in fact a missing data problem. If we define y as the observed outcome, the missing data problem is easily seen by means of a switching equation:

$$y = (1 - w) \cdot y_0 + w \cdot y_1 \tag{2.3}$$

Where y is the observed effect. As it appears in Wooldridge (2002), the problem of estimating (2.1) and (2.2) from a random sample $\{(y^i, w^i)\}^i : i = 1, 2, \cdots N$ becomes the central issue. It is fair to say that great part of the literature on program evaluation, in statistics and economics, deals with the missing data problem. The impossibility of observing (y_0, y_1) for the same individual at the same time forces the analyst to depart from the individual towards the aggregate. So, even though it is impossible to measure the individual impact, nonetheless the average group impact is a feasible measure. It does not come as a surprise that estimators are devised fundamentally to overcome this problem.

In order to have a different perspective on the evaluation problem, we return to the specification in equation (2.3), the switching regression[4]. Let the vector (y_0, y_1) be defined by:

$$\mathbf{E}(y_j|x) = \mu_j(x) = x'\beta_j \;\; j = 0, 1 \tag{2.4}$$

The vector X is a vector of regressors and β_j is a vector of structural parameters. From the definition of conditional expectation and from the tautology $U_j \equiv y - \mu_j(x)$, it is obvious that $\mathbf{E}(U_j|x) = 0$. Hence, substituting (2.4) into (2.3) we obtain the switching regression model:

$$y = (1 - w) \cdot \left[x'\beta_0 + u_0\right] + w \cdot [x'\beta_1 + u_1] \tag{2.5}$$

This can be rearranged to obtain:

$$y = x'\beta_0 + wx \cdot [\beta_1 - \beta_0] + \left[u_0 + w \cdot (u_1 - u_0)\right] \tag{2.6}$$

[4]See Quandt (1972).

Equation (2.6) is an "unusual" regression model. The error term, $u_0 + w \cdot \big(u_1 - u_0\big)$ contains one of the regressors, w. Also, the coefficient of w, $x \cdot [\beta_1 - \beta_0]$, is nothing but the ATE. A simpler framework is helpful to shed more light: assume that the vectors β_0 and β_1 have all components the same except the first ones, β_{10} and β_{00}. Also, let $\beta_{10} - \beta_{00} = \alpha$. With this in mind, a simpler model is:

$$y = x'\beta_0 + w \cdot \alpha + \Big[u_0 + w \cdot \big(u_1 - u_0\big)\Big] \tag{2.7}$$

It has still a non-standard error term. This error is likely to have a non-zero mean: $\mathbf{E}\Big[u_0 + w \cdot \big(u_1 - u_0\big)\Big] = \mathbf{E}[u_1 - u_0|w = 1] \cdot Pr(w = 1)$. If $\mathbf{E}[u_1 - u_0|w = 1] \neq 0$, or put differently, if people choose to participate based on $u_1 - u_0$ or on variables correlated to it, the error term will have a non-zero mean. This is the classical selection bias problem.

Define $\Delta \equiv y_1 - y_0$. A way to estimate ATE_1 by a switching regression is immediate if first we note that:

$$ATE_1 = \mathbf{E}(\Delta|x, w = 1) = \alpha + \mathbf{E}(u_1 - u_0|x, w = 1) \tag{2.8}$$

Using the expression for ATE_1 in equation (2.7):

$$y = x'\beta_0 + w \cdot \alpha + \Big[u_0 + w \cdot \big(u_1 - u_0\big)\Big] + \big\{\alpha + \mathbf{E}(u_1 - u_0|x, w = 1)\big\}$$
$$- \big\{\alpha + \mathbf{E}(u_1 - u_0|x, w = 1)\big\} \tag{2.9}$$

After doing the substitutions, the final equation is:

$$y = x'\beta_0 + w \cdot \big[\alpha + \mathbf{E}(u_1 - u_0|x, w = 1)\big]$$
$$+ \Big[u_0 + w \cdot \Big\{\big(u_1 - u_0\big) - \mathbf{E}(u_1 - u_0|x, w = 1)\Big\}\Big] \tag{2.10}$$

Again, the error term, $\overline{u} = u_0 + w \cdot \Big\{\big(u_1 - u_0\big) - \mathbf{E}(u_1 - u_0|x, w = 1)\Big\}$, is non-standard and the coefficient of w, $\alpha + \mathbf{E}(u_1 - u_0|x, w = 1)$, is the ATE_1. Thus, given equations (2.7) and (2.10) three situations of interest may arise:

- **CASE I** $\Big(u_0 = u_1 = \overline{u}\Big)$

  In this case, $ATE = ATE_1 = \alpha$, equations (2.7) and (2.10) are the same and reduce to $y = x'\beta_0 + w \cdot \alpha + \overline{u}$. This regression model is amenable to conventional econometric techniques. Also, the effect on someone with the same set of characteristics $x^* \in X$ is the same. The key issue is how to eliminate the covariance between w and $\overline{u}$.

- **CASE II** $\Big(u_0 \neq u_1$ but $\mathbf{E}(u_0 - u_1|x, w = 1) = 0\Big)$

  The condition that $\mathbf{E}(u_0 - u_1|x, w = 1) = 0$ means that unobservable variables affecting the gains from the program do not affect the individuals' decision to participate. In this case the same intuition remains valid. Still we have that $ATE = ATE_1 = \alpha$, but people with the same set of observable characteristics, $x^* \in X$ do not necessarily have the same response to the treatment. Besides the issue of the covariance between w and $\overline{u}$, the "error term" is heteroscedastic:

$$\begin{aligned} \mathbf{Var}\big[u_0 + w \cdot (u_1 - u_0)|x, w\big] &= \mathbf{Var}(u_0|x, w) + 2\mathbf{Cov}(u_0, u_1 - u_0|x, w) \cdot w \\ &+ \mathbf{Var}(u_1 - u_0)|x, w) \cdot w \end{aligned} \tag{2.11}$$

- **CASE III** $\Big(u_0 \neq u_1$ and $\mathbf{E}(u_0 - u_1|x, w = 1) \neq 0\Big)$

  This is the worst scenario. The error has non-zero mean and is heteroscedastic. Also, conventional econometrics techniques either do not work or require considerable modifications.

The cases stated before are fundamental to the whole literature on the econometric evaluation of programs. It encompasses all possible situations that someone might face. All different econometrics estimators are necessarily aiming at addressing the issues raised by one of the cases. Before proceeding to the section on different econometric estimators, we discuss a set of measures of the effect of treatments that has been gaining attention. This set is characterized by having a greater concern about the distributive effects of programs not captured by ATE and ATE_1 and it has been mainly developed by James Heckman from University of Chicago and his associates. The next paragraphs draw heavily on Heckman, Lalonde, and Smith (1999).

A weakness of measures such ATE and ATE_1 is that they do not consider that in a democratic society both efficiency and equity issues are relevant in order to evaluate the effect of a program. While these two measures can reflect efficiency gains, they fail to convey anything about equity. Since the existence of a program entails redistribution of income or wealth[5], answers to many important evaluation questions demands the estimation of distributional gains. Heckman, Lalonde, and Smith (1999) provides five measures that might be important to know if equity would be a concern in the evaluation of a program:

1. The proportion of people taking the program who benefit from it:

$$Pr(y_1 > y_0|w = 1) \equiv Pr(\Delta > 0|w = 1); \tag{2.12}$$

[5] Either among participants or between taxpayers and participants.

2. The proportion of the total population benefiting from the program:

$$Pr(y_1 > y_0|w = 1) \cdot Pr(w = 1) \equiv Prob(\Delta > 0|w = 1) \cdot Prob(w = 1); \quad (2.13)$$

3. Selected quantiles of the impact distribution:

$$\inf_{\Delta}\{\Delta : \mathbf{F}(\Delta|w = 1) > q\} \quad (2.14)$$

Where $\mathbf{F}(.)$ is the cumulative distribution function of Δ;

4. The distribution of gains at selected base state values:

$$\mathbf{F}(\Delta|w = 1, y_0 = \overline{y_0}); \quad (2.15)$$

5. The increase in the proportion of outcomes above a certain threshold due to a policy:

$$Pr(y_1 > \overline{y}|w = 1) - Pr(y_0 > \overline{y}|w = 1). \tag{2.16}$$

The importance of these five distributional measures can not be overstated, specially if the program under consideration is large enough to make the stable unit treatment value (SUTVA) assumption unrealistic and the question of indirect effects prominent. Notwithstanding their growing importance, we will focus only on the ATE and ATE_1 effects in the rest of this literature review. This appears to be a sound approach since, as stated in Heckman, Lalonde, and Smith (1999), *"The evaluation problem in its most general form for distributions of outcomes is formidable and is not considered in depth either in this chapter or in the literature ...".*

2.2.2 Social Experiments

Although of relatively recent general application, especially in European countries, the use of social experiments to overcome the evaluation problem appears to be gaining widespread approval since recent years. The main explanation for that seems to be, accordingly to Heckman, Lalonde, and Smith (1999), the simplicity of the analysis and understanding of the model and results. In fact, this feature of social experiments justifies its earlier appearance on this literature review. Hence, non-experimental methods will be the subject of the following subsection. A first necessary step must be the definition of a social experiment[6].

For our purposes, a social experiment is any program such that receiving the treatment is random from the potential participant's perspective. Of course, in the usual program the randomization is performed by the program's managers. The randomization could appear either during the eligibility phase or after, during the treatment phase. Randomization of eligibility is thought to be a better approach than randomization of treatment. However, our focus will be on the latter. From the perspective of the counterfactual framework developed in subsection 2.2.1, a social experiment can be represented by a random vector (y_0^*, y_1^*, w^*, R). The reason for using the superscript star is to contrast the social experiment with the non-experimental set up denoted by (y_0, y_1, w). The two first terms have the same meaning as before. The term w^* has a subtle but crucial different meaning than that of w. Whereas deciding to participate and actually receiving the treatment are the same as having $w = 1$, the meaning of $w^* = 1$ is just that the individual decided to participate in the program. The latter does not necessarily means that the treatment will be delivered. The random variable R represents the randomization and it is defined conditionally on the event $w^* = 1$. So, if $(w^*, R) = (1, 1)$ the individual who has chosen to participate will be allowed to do so, and if $(w^*, R) = (1, 0)$ access to the program will be denied. The set of people with $R = 0$ are called the controls

[6]The term "Social Experiment" is not universally used, though. An alternative term is "Random Experiment". It appears that indeed random experiments are a subset of social experiments. However, we stick to the term social experiments for historical reasons.

and those with $R = 1$ are the treatments. The next assumption is fundamental for the validity of social experiments:

Assumption 1. *There is no randomization bias, or, equivalently:*

$$\mathbf{E}(y_0|x, w = 1) = \mathbf{E}(y_0^*|x, w^* = 1) \text{ and } \mathbf{E}(y_1|x, w = 1) = \mathbf{E}(y_1^*|x, w^* = 1).$$

The interpretation of Assumption 1 is that the implementation of a social experiment will not disrupt the original economic structure where the non-experimental set up would operate.

Suppose that one is interested in estimating the ATE_1. From the social experiment set up, two different kinds of observations can be obtained: $(y_1|w^* = 1, R = 1)$, from those randomized into the program, and $(y_0|w^* = 1, R = 1)$, from those randomized out of the program. Thus, from Assumption 1 and considering the random feature of R, it is clear that social experiments solve the evaluation problem by using observations on those who were randomized out of the program as a proxy for the unobservable quantity $(y_0|w = 1)$, the would-be outcome of non-treatment for those who have received the treatment. More formally, $(y_0|w^* = 1, R = 1) = (y_0|w = 1)$. Making explicit the conditioning on a vector of characteristics X, the desired parameter can be obtained by:

$$ATE_1 = \mathbf{E}(y_1^*|x, w^* = 1, R = 1) - \mathbf{E}(y_0^*|x, w^* = 1, R = 0) \tag{2.17}$$

Hence, the outcome of those randomized into the program and of those randomized out supply the necessary information to calculate the ATE_1. A prevalent wrong conception about the actual mechanism by which randomization operates has been that it makes $\mathbf{E}(u_0|w = 1) = 0$ and $\mathbf{E}(u_1|w = 1) = 0$. Not only this is wrong but also it does not even make $\mathbf{E}(u_1 - u_0|w = 1) = 0$. Randomization simply equates the bias between the groups where $R = 1$ and $R = 0$, $\mathbf{E}(u_1 - u_0|w = 1, R = 1) = \mathbf{E}(u_1 - u_0|w = 1, R = 0)$. Another benefit of randomization often unnoticed is that it also enriches the support of X. Suppose that in the population:

$$Support(X|w = 1) \neq Support(X|w = 0) \tag{2.18}$$

Hence, for those individuals in the treatment group who can not be matched with controls with the same set of observed characteristics, or vice-versa, $\mathbf{E}(\Delta|x, w = 1)$ can not be identified. Randomization enriches the support by making:

$$Support(X|w = 1, R = 1) = Support(X|w = 0, R = 0) \tag{2.19}$$

A different perspective to understand the effect of randomization is to place it in a context of instrumental variable estimation, as shown in Heckman (1996). In the following, the variable R works as an instrumental variable. Assuming no randomization bias and without loss of generality equating w and w^*, define:

$$\widetilde{y} = y|(x, w = 1) \tag{2.20}$$

$$\widetilde{u_0} = u_0|(x, w = 1) \tag{2.21}$$

$$\widetilde{u_1} = u_1|(x, w = 1) \tag{2.22}$$

These are random variables that represent the realizations of y, u_0 and u_1 conditional on $w = 1$ and x. Since R and $\widetilde{y}$ are independent, equation (2.6) can be rewritten as:

$$\widetilde{y} = x'\beta_0 + Rx \cdot [\beta_1 - \beta_0] + \left[\widetilde{u_0} + R \cdot \left(\widetilde{u_1} - \widetilde{u_0}\right)\right] \tag{2.23}$$

Adding $R \cdot \left\{\mathbf{E}(u_1|x, w = 1) - \mathbf{E}(u_1|x, w = 1)\right\} + (1 - R) \cdot \left\{\mathbf{E}(u_0|x, w = 1) - \mathbf{E}(u_0|x, w = 1)\right\}$ to equation (2.23) and rearranging:

$$\widetilde{y} = \left\{x'\beta_0 + \mathbf{E}(u_0|x, w = 1)\right\} + R \cdot \left\{x\beta_1 - x\beta_0 + \mathbf{E}(u_1 - u_0|x, w = 1)\right\}$$
$$+ \left\{\widetilde{u_0} - \mathbf{E}(u_0|x, w = 1) + R \cdot \left[\widetilde{u_1} - \mathbf{E}(u_1|x, w = 1) - \widetilde{u_0} + \mathbf{E}(u_0|x, w = 1)\right]\right\} \tag{2.24}$$

Using equation (2.24), noting that $x\beta_1 - x\beta_0 + \mathbf{E}(u_1 - u_0|x, w = 1)$ is the ATE_1 and defining :

$$\varphi(x) = x'\beta_0 + \mathbf{E}(u_0|x, w = 1) \tag{2.25}$$

$$\widehat{u_0} = \widetilde{u_0} - \mathbf{E}(u_0|x, w = 1) \tag{2.26}$$

$$\widehat{u_1} = \widetilde{u_1} - \mathbf{E}(u_1|x, w = 1) \tag{2.27}$$

It follows that:

$$\widetilde{y} = \varphi(x) + ATE_1(x) \cdot R + \left(\widehat{u_0} + R \cdot [\widehat{u_1} - \widehat{u_0}]\right) \tag{2.28}$$

Conditional on X the equation above is a simple regression model, with intercept $\varphi(x)$. The coefficient of R, say $ATE_1(x)$, is the desired treatment on the treated effect. Clearly $\mathbf{E}\big(\widehat{u_0} + R \cdot [\widehat{u_1} - \widehat{u_0}]|R\big) = 0$ and $\mathbf{E}\big(\widehat{u_0} + R \cdot [\widehat{u_1} - \widehat{u_0}]\big) = 0$. Hence, randomization acts as an instrumental variable.

Note, however, that while it is possible to consistently estimate the $ATE_1(x)$ in equation (2.28), the "deep" structural parameters, (β_0, β_1), can not be identified. Thus, the data produced by social experiments, although solving the evaluation problem, do not estimate parameters usually of interest to economists[7].

It is important not to overestimate the strengths of social experiments. Besides the assumptions stated before, its properties depend crucially, as any other non-experimental method, on the absence of dropout and group substitution. The latter means the attempt by controls of receiving training in other programs substitute to the program under evaluation. Substitution clearly bias the estimated parameters. Also, social experiments still fall short beyond estimating average effects. Strong assumptions are required if distributions are to be estimated. Finally, even under a social experiment there are some important questions that require the use of non-experimental methods. A good example is contained in Ham and LaLonde (1996). They noted that to calculate the effect of training of participants in the National Supported Work(NSW) program on their posterior unemployment(search) and employment duration non-experimental methods should be used. The reason is simple: even though treatments are randomly selected, the search time is subject to time dependence and/or unobserved heterogeneity. Hence,

[7] One way to overcome this identification problem is to rely in a different experiment. If the interest rests upon estimating $y = g(x) + u$, and x is endogenous, x can be independently varied as in the negative income tax experiments. See, Heckman (1992).

the sample to be used to compare the posterior employment duration is clearly non-random. This also would be the case if wages were the outcome under measure. Only wages for those employed would be observable and again the only available sample would be non-random. Nonetheless, the situation would be even worse if the selection into the program was not random. Next section surveys the huge amount of methods available to evaluate non-experimental programs.

2.2.3 Non-experimental Programs

When randomization is not available different ways to deal with the evaluation problem are necessary. Methods to evaluate the effect of treatment in a non-experimental setup has been developed since the pioneering works of Rubin (1974) and Rosenbaum and Rubin (1983). This literature is huge, so we tentatively survey it by closely following Wooldridge (2002). As a first step we restate **the evaluation problem** and develop an extended stochastic model.

The evaluation problem as analyzed before is a missing data problem, or, the impossibility to observe for the same person, at the same time, both y_0 and y_1. Without assuming randomization, the first set of non-experimental methods are based on an assumption due to Rosenbaum and Rubin (1983): ignorability of treatment. Since regressors play a key role now, let us define our population model as the random vector (y_0, y_1, w, x). Define stochastic independence by the symbol $\perp$. Thus the ignorability of treatment assumption is:

Assumption 2. *The vector of potential outcomes is independent of participation in the program conditional on a set of regressors, or:*

$(y_0, y_1) \perp w|x.$

For example, if we have enough observed characteristics (by both the individual and econometrician) of the individual and the program it might be the case that conditioning on those observables the choice to either participate or not does not depend on the potential gains. Sometimes a weaker assumption suffices, conditional mean independence:

Assumption 3. *The vector of potential outcomes is mean independent of participation in the program conditional on a set of regressors, or:*

$\mathbf{E}(y_0|x, w) = \mathbf{E}(y_0|x)$ *and* $\mathbf{E}(y_1|x, w) = \mathbf{E}(y_1|x)$.

Assumptions 2 and 3 mean that by conditioning on enough regressors it is possible to make potential outcomes independent of the treatment indicator. The way the evaluation problem is solved is straightforward. The missing part of the ATE_1, say $\mathbf{E}(y_0|x, w = 1)$, is easily obtained by noting that $\mathbf{E}(y_0|x, w = 1) = \mathbf{E}(y_0|x, w = 0) = \mathbf{E}(y_0|x)$. So, non-treated people supplies the necessary observations. The question of how many regressors is enough and how to choose them is still open[8]. Some authors such as Heckman and Roob (1985) and Moffitt (1996) call assumptions like (2) and (3) selection on observables.

[8]If w is a deterministic function of x the assumptions clearly hold. For instances, if a program participation is granted for those who earn no more than a threshold.

It is worth mentioning that assuming selection on observables requires no additional assumptions, such as functional forms, in order to identify both ATE and ATE_1.

There are two major drawbacks of conditioning on observables to estimate average effects. First, it is very likely that after finding a convenient vector X to make either assumption 2 or 3 work some realization of X, say x^*, have very few observations in the sample to make estimation at this point not reliable. Second, the dimension of X can be an impediment to estimation, specially if non-parametric methods are to be used. An alternative to condition on the entire set of regressors X was proposed by Rosenbaum and Rubin (1983) and it consists of conditioning on the propensity score:

Definition 1. *The Propensity Score is the probability of selection conditional on a set of regressors:* $\mathbf{p}(x) = Pr(w = 1|x)$.

If we impose the next assumption, it turns out that both ATE and ATE_1 can be written in terms of the propensity score:

Assumption 4. *The propensity score is an interior point of the unit interval:* $\mathbf{p}(x) \in (0, 1)$.

It follows that:

$$ATE = \mathbf{E}\left(\frac{[w - \mathbf{p}(x)]y}{\mathbf{p}(x)[1 - \mathbf{p}(x)]}\right) \tag{2.29}$$

$$ATE_1 = \mathbf{E}\left(\frac{[w - \mathbf{p}(x)]y}{[1 - \mathbf{p}(x)]}\right) \cdot \frac{1}{Pr(w = 1)} \tag{2.30}$$

The key identifying propriety of the propensity score is due to the fact that Assumptions (2) and (4) imply that $(y_0, y_1) \perp w|\mathbf{p}(x)$. The proof is simple:

Proof:

Under ignorability of treatment (assumption (2)), $\mathbf{E}(w|y_0, y_1, x) = \mathbf{E}(w|x) = \mathbf{p}(x)$.

Also, $\mathbf{p}(x) = \mathbf{E}(w|\mathbf{p}(x))$ by iterated expectations. By applying again the law of iterated expectations we have:

$$\begin{aligned}
\mathbf{E}(w|y_0, y_1, \mathbf{p}(x)) &= \mathbf{E}\big[\mathbf{E}(w|y_0, y_1, x)|y_0, y_1, \mathbf{p}(x)\big] \\
&= \mathbf{E}(\mathbf{p}(x)|y_0, y_1, \mathbf{p}(x)) \\
&= \mathbf{p}(x)
\end{aligned}$$

Q.E.D

Estimation by conditioning on the propensity score reduces the dimension of the conditioning vector to one. Of course, if $\mathbf{p}(x)$ is not know it must be estimated by parametric on non-parametric methods. A recent nice feature of models based on the estimated propensity score, $\widehat{\mathbf{p}}(x)$, is that their estimator can achieve very small asymptotic variances, as shown in Hirano, Imbens, and Ridder (2000). A very different way to use the propensity score to estimate treatment effects is the matching approach suggested by Rosenbaum and Rubin (1983).

The matching approach "matches" people with the same propensity score[9] who belongs to both control and treatment groups. In light of the preceding discussion, $\mathbf{E}[y_1 - y_0|\mathbf{p}(x)]$ can be estimated by the sample average, and by iterated expectations, the $ATE = \mathbf{E}[y_1 - y_0]$ can be obtained by averaging across the distribution of the propensity score. Some recent examples of this approach are Heckman, Ichimura, and Todd (1997) and Angrist (1998). A likely situation in the evaluation of programs is that participation might be correlated with unobservables. If this is true, methods that condition only on observables are not helpful. Fortunately, an instrumental variables approach (IV) is a sound strategy for those cases.

Since the use of IV estimators in the context of evaluation of programs were discussed in Section (2.2.2), the following discussing will be short. To motivate the second set of non-experimental estimators, equation (2.7) is used as the benchmark model[10]:

---

[9] In practice each cell is composed of a interval of propensity scores values, $cell_j = (\underline{p}^j, \overline{p}^j)$.

[10] It is assumed for the sake of simplicity that $u_0 = u_1 = \overline{u}$.

$$y = x'\beta_0 + w \cdot \alpha + \overline{u} \tag{2.31}$$

Where $\alpha = ATE$ and it is not assumed that $\overline{u}$ is mean independent of w given x, or, $\mathbf{E}(\overline{u}|w,x) = \mathbf{E}(\overline{u}|x)$. So, w is endogenous and the model depicted in equation (2.31) is also known as the dummy endogenous variable model for obvious reason. The fundamental problem is to find good instruments. An often used instrument is randomization of eligibility. Let $\mathbf{z}$ be a dummy random variable indicating if a person is eligible or not to participate in a program. Variable $\mathbf{z}$ is clearly a good candidate for an instrument: it is definitely independent of $\overline{u}$ and clearly correlated with w. After choosing a good instrument, equation (2.31) can consistently estimate ATE. It appears, though, that "selection on unobservables" models have been receiving diminishing attention. The increasingly detailed data sets available from recent programs appear to leave less to be explained by unobservables[11]. However, the role of instrumental variables would be at center stage when discussing the final non-experimental estimator: the local average treatment effect (LATE) of Imbens and Angrist (1994).

In order to motivate the LATE estimator, we begin by describing a fictitious situation. Suppose we are trying to asses the effect of a training program run by the

[11] It is fair to say that this is true if we do not compare the non-experimental program to a would-be social experiment. Although matching the right people is more important than selection on unobservables, there is a considerable bias after controling for observables when a ideal social experiment is used as a benchmark. This point was made very clear in Heckman, Ichimura, and Todd (1997).

government on some labor market outcome. We are sure that there is selection bias and it operates through unobservables. For simplicity, it is assumed that a correctly specified model such as equation (2.31) without assuming common treatment effect represents the program:

$$y = x'\beta_0 + w \cdot [x'\beta_1 - x'\beta_0] + \overline{u} \tag{2.32}$$

As discussed, a sound strategy would be to find a good instrument. Assume, for illustrative purposes, a good instrument is the distance from the treatment center, Z. Assume that the participation indicator w is a function of Z, $w = h(z)$. The observable sample is $\{y_i, z_i, w_i\}$, for $i = 1, 2 \cdots N$. Besides the key issue of estimating the treatment effects, the government is trying to answer the following question: How the estimates of the program's impact will change if the location of the treatment center is altered? First, note that is a legitimate question because distance from the treatment center influences program participation. Second, from a theoretical perspective what is being asked is the relevance of the instrumental variable approach to estimate different effects. It turns out that the answer is that, unless very restrictive assumptions are imposed, an IV estimator can not identify the effect that the government wants to know. To see this, note that the desired effect is indeed the local average treatment effect, or, $LATE = \mathbf{E}[y|z'] - \mathbf{E}[y|z'']$, and that:

$$\mathbf{E}[y|z'] - \mathbf{E}[y|z''] = \mathbf{E}[h(z') \cdot x'\beta_1 - (1 - h(z')) \cdot x'\beta_0 | z'] \\ - \mathbf{E}[h(z'') \cdot x'\beta_1 - (1 - h(z'')) \cdot x'\beta_0 | z''] \tag{2.33}$$

Since, Z is an instrumental variable:

$$\mathbf{E}[y|z'] - \mathbf{E}[y|z''] = \mathbf{E}\big[(h(z') - h(z')) \cdot (x'\beta_1 - x'\beta_0)\big] \\ = Pr[h(z') - h(z'') = 1] \cdot \mathbf{E}[x'\beta_1 - x'\beta_0 | h(z') - h(z'') = 1] \\ - Pr[h(z') - h(z'') = -1] \cdot \mathbf{E}[x'\beta_1 - x'\beta_0 | h(z') - h(z'') = -1] \tag{2.34}$$

The term $\mathbf{E}[x'\beta_1 - x'\beta_0 | h(z') - h(z'') = 1]$ is the treatment effect for those who switched from program participation, $h(z') = 1$, to non participation, $h(z'') = 1$, when the instrumental variable changed from z' to z''. Analogously, the term $\mathbf{E}[x'\beta_1 - x'\beta_0 | h(z') - h(z'') = -1]$ represents the treatment effect for those who switched from non participation to participation. The identification problem arises because even if the effect of the treatment is positive for every participant, equation (2.34) shows that $\mathbf{E}[y|z'] - \mathbf{E}[y|z'']$ can be zero, positive or negative. The government's question can not be answered without additional restrictions. An intuitive reason for this is, accordingly to Imbens and

Angrist (1994), that *" ... the treatment effect for those who shift from non participation to participation when Z is switched from $[\mathbf{z}']$ to $[\mathbf{z}'']$ can be cancelled out by the treatment effect of those who shift from participation to non participation"*. Imbens and Angrist (1994) solves the identification problem by assuming a monotonicity condition:

Assumption 5. *For all $z', z'' \in Support(Z)$, either $h(z') \geqq h(z'')$ for all individuals or $h(z') \leqq h(z'')$ for all individuals.*

Hence, with Assumption (5), Imbens and Angrist (1994) shows that the LATE is identified for $z' \neq z''$. We omit the proof here and suggest the original paper or Wooldridge (2002), Chapter 18, to the more interested reader.

The literature about econometric evaluation of programs has developed into various directions in the last decade. Most recent developments are programs with multi-treatments and/or vector of outcomes, continuous treatment, programs with indirect effects, heterogeneity in program's impact, structural models of program's participation, and programs in a duration analysis context. For an excellent account of those developments, see Heckman, Lalonde, and Smith (1999) and the extensive literature cited there are the recommended starting points. The next section surveys the literature on program evaluation in a context of duration analysis. Although duration analysis is a well established topic in the econometric literature[12], models that try to address evaluation of programs in that context is scarce.

[12] For instance, see Lancaster (1990).

2.2.4 Treatment Effect and Duration Analysis

For our purposes, a program in a duration analysis context is defined as any program such that either the vector of outcomes (y_0, y_1) or the time of receiving the treatment[13] w or both are duration variables. Not surprisingly, to evaluate those type of programs, the econometrician faces the challenges posed by both fields of evaluation of programs and duration analysis. To organize the discussion of this subsection, we call programs with an outcome described by a duration variable, duration outcomes (DO), and call those with a stochastic time of occurrence of $w = 1$, duration treatment (DT). A third possibility is a program having both features[14]: we call them duration outcome and treatment (DOT).

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 would follow the first duration, there would be no problem if 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 ATE. Hence, the selection problem arises if any variable whose occurrence depends on the

[13] More precisely, we are interested in situations where $w = 1$ for everybody, but the timing of the treatment is random, however.

[14] As far as we known, 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.

realization of the first duration is the object under evaluation, regardless if it a duration variable or not. The following example should help understand the situation.

Suppose a program offers training with the intent of both shortening the time unemployed and raise 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 4 possible realizations: controls still unemployed, (CSU), controls employed, (CE), treatments still unemployed, (TSU), treatment employed, (TE). The interest is on measuring the effect of the program on the initial wages. An analogous way to estimate ATE appears to be 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 $t = t^*$, will have, on average, a much higher wage than the sample of controls. As a consequence, the ATE is overestimated. The reason for this bias is clarified in Ham and LaLonde (1996) : *"[the outcome of interest] is missing for some individuals and whether is missing depends on an individuals' experimental status and unobservables ... "*.

The key message from Ham and LaLonde (1996)' paper[15] is that if one wants to analyze programs with duration outcomes, even with social experiments 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, Ham, and Lalonde (1997). Both papers try to evaluate experimentally two DO programs, the National Supported Work Demonstration (NSWD) and Job Training Partnership Act (JTPA), that posses the following (sequence) of outcomes shown in Figure 2.1.

Fig. 2.1. Sequencing at the JTPA

$$\boxed{\text{TREATMENT}} \longrightarrow \boxed{\text{SEARCH}} \longrightarrow \boxed{\text{EMPLOYMENT}} \longrightarrow \boxed{\text{UNEMPLOYMENT}}$$

They are interested in measuring the effect of treatment on the duration of employment and posterior unemployment. Since search time is a duration variable, their programs are DO. Note that for those programs we have duration variables following the initial duration variable. This makes their analysis even more difficult than the stylized example of the wage outcome. Finally, both authors developed some intuition about the importance of dynamic selection at different set ups. After some cautions, they relate the

[15]It is worth mentioning that DO programs have been explored before in the econometric literature, for instance, in Card and Sullivan (1988) and Meyer (1990). Those papers did not address the key dynamic selection problem, though.

importance of dynamic selection to the length of time that the program keep individuals out of the labor market. So, in Eberwein, Ham, and Lalonde (1997) they conclude by stating that the longer a program keep an individual out of the labor market the more important will be the issue of dynamic selectivity 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 set up of even more recent interest is the econometric evaluation of DT programs.

In DT programs the timing of occurrence of the treatment is a 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. We are interested in measuring the effect of this training on the time of finding a job[16]. First, note that we are interested not only in the occurrence of an event but also on the time of this occurrence. This is in contrast with the literature of program evaluation that is concerned only with the binary indicator w. Second, as it will be clear, 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

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

Lillard (1993) who estimates the effect of having a child on the duration of marriage. In that paper, the time 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 death time of people. A labor market application appears in van den Berg, van den Klaauw, and van Ours (1998). The development of identification of a specific class of models to deal with DT programs is considered in Abbring and den Berg (2000a). In fact, their paper is the point of departure of our econometric model of the ESEO program. Since a full development of a model analogous to those appearing in Abbring and den Berg (2000a) appears in Section 2.5 of our paper, we conclude this subsection by briefly commenting on a final issue.

The identification issue that arises in the context of DT programs is the possible lack of distinction between the effect of treatment and dynamic selectivity. To see this suppose that we have an initial group of people that can be characterized by a vector of observable variables X and a vector of unobservable variables V. Hence, the vector (T_{basic}, T_{treat}) is a function of (X, V). Now consider the problem of evaluating the effect of the treatment using a sample of individuals who have been treated at date $T_{treat} = t^*$. This sample is not a random sample from the population, however. There are two reasons why this is indeed the case. First, since T_{treat} is a function of V the distribution of those with $T_{treat} = t^*$ is not the same as the distribution on the original population. This is just the classical selection problem in duration analysis. Second, to observe someone with $T_{treat} = t^*$ is necessary that $T_{basic} \geqslant t^*$, so the distribution of T_{basic} differs between sample and population. Again, this is the classical selection problem in duration analysis. Thus, and individual with treatment at $T_{treat} = t^*$ could have an lower value of T_{basic} because either the treatment is effective or he/she would have a lower value anyway due to unobservables or both.

It turns out to be that accounting for the timing of occurrence of the treatment facilitates identification. Following Abbring and den Berg (2000a), an intuitive explanation for that relies on the fact that the time of occurrence provides extra information when compared only to the knowledge that w is 0 or 1. Also, the same individual is both a control and treatment because to estimate the treatment effect, observations before the realization of the treatment are compared to observations after the realization of the treatment[17]. The effect of the treatment can be distinguished from the effect of unobservable variables and identification results. The next Section gives a detailed description of the institutional features of the ESEO program.

2.3 Characterizing the Program

2.3.1 Antecedents

During the last decades, sociologists and economists have been devising programs to ease the difficult transition faced by ex-offenders during the period of time between release and reintegration into society. As experience has accumulated, a fundamental goal to a complete reintegration turned out to be job placement. A good job would be necessary not only to provide the basic needs for survival in the short run but also as

[17] Observe, that censored observations are also informative as long as the censoring occurs after the realization of the treatment.

a key element to secure self-esteem, security and sense of integration in the society as whole. Hence, sociological and economic theory have provided enough justification for the existence of employment services programs for ex-offenders.

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 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 Mylkman, Timrots, Peyser, Toborg, Yezer, Carpenter, and Landson (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 have resulted in the appearance of programs that had a strong preoccupation with the post-placement of their clients. These programs have designed follow-up strategies to overcome the major criticism of past experience. Among various programs, three deserve recognition for both being successful and having similar structures: the Comprehensive Offender Resource

System, in Boston, the Safer Foundation, in Chicago and Project JOVE, in San Diego. Not surprisingly, the U.S. Department of Justice saw this as a opportunity for assessing the efficacy of employment services programs that contained a follow-up component. Then, in 1985 the Department of Justice funded a research performed by the Lazar Institute from McLean, VA. The next subsection describes the institutional details common to all three programs[18]. For easy of exposition this common "program" is referred to as the Employment Services for Ex-offenders (ESEO) program.

2.3.2 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. We postpone the two first aspects until Section 2.4, where the details about the available data set is discussed , since we believe to be a more appropriate place. Hence, in the ESEO program[19], after being assigned to either the control or treatment group[20] the clients step inside the intake unit, where they received initial orientation, screening and evaluation by an intake counsellor. While still in this first phase, to secure survival up to the job search phase, the intake counsellor offered minimal assistance services such as food, transportation, clothing and etc.

---

[18] Of course, no program was identical to the others. However, specific attributes were not relevant to deserve separate analysis.

[19] We closely follow Mylkman, Timrots, Peyser, Toborg, Yezer, Carpenter, and Landson (1985).

[20] Unfortunately, there is know information whether clients knew their treatment status. As it will be clear later on, in the context of DT programs the state of knowledge about treatment status could help identified other parameters of interest besides the effect of treatment.

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 like 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 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, Lalonde, and Smith (1999). The job search is the last phase in the ESEO program that is offered equally for both controls and treatments. It is fair to say that there is difference (observable) between the job search of controls and treatments. While the former is helped only by the job developer the latter has both the job developer and the follow-up counsellor help. Nevertheless, at this phase, the follow-up counsellor helps only tangentially, restricting himself/herself to weekly meetings. Hence, for our purposes both controls and treatments receive the same service. The difference begins upon placement. Figure 2.2 helps understand the sequence of services.

Fig. 2.2. Sequence of Services on the ESEO Program

INTAKE

$\downarrow$

SEARCH SKILLS

$\downarrow$

SEARCH ASSISTANCE

$\Downarrow$

FOLLOW-UP (TREATMENTS ONLY)

Controls were not helped after placement whereas treatments started receiving follow-up help just after the employment relation starts[21]. The follow-up special services consisted basically of crisis intervention, counselling 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 of each individual program, one should consult Timrots (1985), Mylkman, Timrots, Peyser, Toborg, Yezer, Carpenter, and Landson (1985) and Milkman (2001).

There are basically two ways to measure recidivism outcomes in the ESEO program : through count data or duration data. Detailed number of arrests from date of released to the end of the program for all clients was gathered in the respective state police departments. That was the data used at the original evaluation made by Mylkman, Timrots, Peyser, Toborg, Yezer, Carpenter, and Landson (1985). Also, data on the first arrest after release is available. Indeed, to conform to our proposed model, duration data is the necessary choice. Since part of the motivation of our paper is to reevaluate the ESEO program, it is worth describing the general findings of the first econometric evaluation. Three main results emerged from their analysis:

- In general, the experiment did not demonstrate that follow-up services decreased the chances of long-term recidivism;
- For some programs, specially the Safer Foundation in Chicago, and some individuals the follow-up services appears to lower recidivism rates;

[21] Placement consisted basically of unsubsidized jobs, although few placements were in public jobs or skill training.

- Some characteristics, such as drug use, long criminal records have a strong negative effect on recidivism, i. e., higher drug consumption or lengthier past criminal records anticipates recidivism; and others, such as older age and married status, have a positive effect in recidivism, i. e., they postpone recidivism.

  Although the efficacy of employment services with follow-up is still under debate, a different model that captures better the nuances on the program could be an important contribution to either corroborate the results of Milkman and associates or to counterpoint their views. We think that a considerable source of misspecification of their models originates from the fact of not considering the timing of the treatment, in their context the time of placement. Thus, reevaluating the ESEO program by modelling it as a duration treatment program seems to be worth pursuing. In this case the basic duration will be the time of recidivism and the treatment duration will be the time necessary to be placed. Before we develop the econometric model, Section 2.4 discusses the available data set.

2.4 The ESEO Program Data Set

The data set used in our present analysis comes from the Inter-University Consortium for Political and Social Research, henceforth ICPSR, under the study number 8619. This study was one of the first to try to assess the impact of programs of reemployment with post-placement support for ex-convicted. As a reaction to the accumulating evidence not entirely favorable to the traditional type of reemployment programs, the National Institute of Justice (NIJ) sponsored a controlled experiment to test the hypothesis that a careful and intensive post-placement follow-up would decrease the level of recidivism.

As noticed by Mylkman, Timrots, Peyser, Toborg, Yezer, Carpenter, and Landson (1985), the traditional type of employment services programs severs contact with the ex-convicted as soon as he or she gets a placement. If any type of follow-up occurs it is limited to only telephone calls to employers to check the employment status of the placed. There is almost no time devoted to crisis intervention or adjustment counselling, which many experts believed to be a crucial step on any attempt to fully reintegrate the ex-convicted into the society. Hence, the NJS through the Lazar Institute designed a experiment to compare the effect of "normal" programs and "special" programs[22] of reemployment. For this purpose, three well established programs were chosen: the Comprehensive Offender Resource System (COERS) in Boston, the Safer Foundation in Chicago, and the Project JOVE in San Diego.

The ESEO data set is comprised 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 ICPSR only made available 1074 usable observations: 325 in Boston, 489 in Chicago and 260 in San Diego. 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 drop-out, features of placement (wage, number of hours, match quality), and so on;

[22] Where "special" programs are "normal" programs with an additional, more structured follow-up component.

- **Outcome variables**: number of arrests, date of first arrest, self-reported arrests for placed people only, and so on.

Before we proceed describing with more detail the available data set, it is important to pay attention to an important message from the literature on evaluating program's effects. As emphasized in Heckman, Lalonde, and Smith (1999), *"One of the most important lessons from the literature on evaluating social programs is that choices made by evaluators regarding their data sources, the composition of their comparisons groups, and the specification of their econometric models have important impacts on the estimated effects ..."*. So, whenever it is possible and we judge necessary, the data set description will be followed by an analysis of the circumstances that surrounded its choice as well as the impact of that choice in the whole model.

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[23]:

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 of program participation;

3. Participants exhibited a pattern of income-producing offenses.

[23] For institutional details about the ESEO program we closely followed the only 2 available published documents, i. e., Mylkman, Timrots, Peyser, Toborg, Yezer, Carpenter, and Landson (1985) and Timrots (1985) .

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

Given our initial sample, the individuals were randomly assigned to either the treatment or control group[24]. Controls receive the standard services and Treatments received, in addition to that, emotional support and advocacy during the follow-up period of 180 days after placement. As it will be clear in the next section, we are going to use a duration model to estimate the impact of the ESEO. Hence, two durations are of great importance: T_s, time spent searching a job and T_c, recidivism time. These two variables are grouped, however. As shown in Appendix B.2, the intervals length varies both within and between durations. Grouped data present no obstacles for the estimation of duration model. Yet, the variance of estimated parameters is obviously higher and identification can be hampered.

The choice of regressors is guided by both T_s and T_c. Our point of departure to choose a set of regressors for recidivism is Schmidt and Witte (1988). However, we also pay close attention to the criminologic literature in recidivism, for instance Gendreau, Little, and Goggin (1996). The regressors used in Schmidt and Witte (1988) are basically age at release, time served for the sample sentence, sex, education, marital status, race,

[24]This randomization will be fundamental for the econometric model.

drug use, supervision status, and dummies that characterize the type of recidivism. The modern literature on criminal recidivism distinguishes two clear groups of recidivism predictors: static factors and dynamic factors. Static factors are variables that belongs to the prisoner past and can not be changed, such as age, sex, history of convictions. Dynamic factors or also known as criminogenic needs, can change and should be a target of the treatment. For example, values, behaviors , and antisocial cognitions. However, as argued by Gendreau, Little, and Goggin (1996), while the use of static factors presents no disagreement the use of dynamic factors is not a consensus in the profession. Hence, we use only static variables.

Different from the literature on recidivism, the literature on unemployment (search) duration has been refined since the 70's. Nowadays, it has a status of a complete theory of unemployment, as it appears in Pissarides (2000). Its empirical contents has been developed since the late 70's and this first wave of empiricism is characterized for 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 in such models. An updated account of that appears in van den Berg (1999). There has been also studies close to ours that try to measure the effect of programs in a context of a model of unemployment(search) duration. For instances, Abbring, van den Berg, and van Ours (1997), Eberwein, Ham, and Lalonde (1997) and van den Berg, van der Klaauw, and van Ours (1998).

As a consequence of those studies, a set of important regressors has been singled out. This set is composed basically of schooling, sex, age, and race. Together with the regressors related to recidivism this form our large set of regressors used in our model. It is worth noting that given the identification requirements of our model, sometimes some exclusions and inclusions restrictions must be imposed. The complete set of regressors appears below and a summary of descriptive statistics appears in Tables 2.1 and 2.2, for controls and treatments, respectively.

Table 2.1. 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	1.0242	0.4529
AGE	1074	16	59	27.4497	6.5739
SANDIEGO	1074	0	1	0.2420	0.4285
CHICAGO	1074	0	1	0.4553	0.4982
MOSUNEMP	1074	0	97	9.3791	19.5649
AGEFIRST	1074	6	44	16.4543	4.5406

Given the significant number of occurrence of attrition, we show two tables. Table 2.1 brings all variables described, except **SEARCH**. This is due to the fact that the attrition occurred during the search period and there is no available data on this censored duration, only the code 88 is available in the original data set. However, such attrition, as pointed before, does not imply censoring on the crime duration. Thus the variable **CRIME** is listed in Table 2.1. To asses both durations for those who have not suffered from attrition, Table 2.2 lists both durations conditional on **ATTRITION** = 0. The list of all variables appears next.

Table 2.2. Durations for ATTRITION = 0

Variable	**Observations**	**Min**	**Max**	**Mean**	**Stdev**
SEARCH	503	1	4	1.5646	0.6644
CRIME	503	1	5	3.3399	1.0008

- **ATTRITION**: Indicator for attrition status. **ATTRITION** = 1 means the individual is either a "no show" or a "drop-out", **ATTRITION** = 0 otherwise;
- **SEARCH**: Discrete variable indicating which interval[25] the search duration belongs to. **SEARCH** = {1,2,3 or 4};
- **CRIME**: Discrete variable indicating which interval[26] the recidivism belongs to. **CRIME** = {1,2,3,4 or 5 };
- **GROUP**: Indicator for group participation. **GROUP** = 0 means control, **GROUP** = 1 means treatment;
- **DRUG**: Indicator for the use of drugs during the last 5 years. DRUG = 0 means no use, DRUG = 1 otherwise;
- **RACE**: Indicator for race. RACE = 0 means white, RACE = 1 means non-white;
- **SEX**: Indicator for 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;

[25] See Appendix B.2.
[26] See Appendix B.2.

- **AGE**: Age of ex-convicted in years;
- **SANDIEGO**: Indicator for city. SANDIEGO = 1 means San Diego, SANDIEGO = 0 means either Chicago or Boston;
- **CHICAGO**: Indicator for city. CHICAGO = 1 means Chicago, CHICAGO = 0 means either San Diego or Boston;
- **AGEFIRST**: Age at first arrest in years.

Next section describes the econometric model built to estimate the effect of follow-up counselling in the recidivism behavior of ex-convicts.

2.5 An Econometric Model of the ESEO Program

2.5.1 Defining a Program

As mentioned before, a distinguishing feature of our program is the fact that the initial date of the treatment depends on the time of search and on the event of placement. Also, one of the outcomes used as a benchmark in evaluating the efficacy of the program, say, recidivism time, is a duration variable whose starting date coincides with the starting date of the search phase. Besides that, our observations are right censored, which is a common feature of duration data sets. To our knowledge, Abbring and den Berg (2000a) is one of the first attempts of modelling treatment effects in a context of duration analysis that rigourously discusses nonparametric identification. Nonetheless, the effect identified by those authors is not the same treatment effect used in the literature of econometric program evaluation. One of the reasons for this stems from the fact that Abbring and den Berg (2000a) lacks a control group for comparison purposes, as it is traditionally present in evaluation studies.

A fundamental assumption in the literature of program evaluation, e.g., Heckman, Lalonde, and Smith (1999), is the existence of a control group. People belonging to this group do not receive the treatment and, assuming the absence of general equilibrium effects, they represent a situation of non existence of the program. Consequently, this group serves as an indispensable benchmark for evaluating the program. While in Abbring and den Berg (2000a) there is uncertainty related to the treatment, this uncertainty is restricted to the timing only. By assuming that the random variable representing time until treatment is non defective, they rule out by assumption the existence of a control group. So, without a control group it is not possible to calculate the gain or loss of implementing the program and therefore, the program can not be evaluated.

The previous critique should not be taken to mean that the model in Abbring and den Berg (2000a) is unimportant. By the contrary, we believe their framework is a very nice starting point available. However, the very definition of what a program is in that context should be revised[27]. Hence, we start by defining precisely the program in terms of a random vector.

Definition 2. *A Program is a random vector* $\mathbf{P} = (\mathcal{T}, I_{sel}, I_{tr}, \mathcal{X}, \mathcal{U}, \Gamma, \overline{T})$. *The vector* $\mathcal{T} = (T_s, T_a, T_c)$, *represents basic durations. The subscripts* s, a *and* c *represents the duration of search, attrition and recidivism, respectively. The indicator* I_{gr} *represents*

[27] It is important to outline that Abbring and den Berg (2000a) do not claim that their model should be taken as a model of program evaluation. They make clear, however, that they want to identify the effect of a “treatment”.

group assignment, and I_{tr} *is an indicator of treatment status.* $\mathcal{X}$ *is a vector of exogenous variables.* $\mathcal{V} = (V_s, V_a, V_c)$ *is a vector of unobserved heterogeneity.* Γ *is a vector of structural parameters and the random variable* $\overline{T}$ *is the censoring time.*

Although not all features of the program will be analyzed, it is instructive to describe the many possible details susceptible of analysis. Upon being selected to the treatment group, an individual simultaneously receives search help and, since he is free, he can commit a crime. The first duration is represented by the random variable T_s and the latter duration is represented by the latent duration variable T_c. The impact of the treatment (follow up) is expected to change some features of T_c. For control individuals the logic is the same except that they do not have the follow up treatment. Both groups can expect the occurrence of attrition. Attrition time can be modelled as an latent duration variable, T_a. The probability of being a treatment or control is respectively $Prob[I_{gr} = 1]$ and $1 - Prob[I_{gr} = 1]$. Finally, there are the no shows and the drop outs. The probability of being a no show is $Prob[NS = 1]$ and the probability of dropping out is $Prob[DO = 1]$. It is worth noting that the dropping out process could be modelled by means of T_a, as in van den Berg, Lindeboom, and Ridder (1994) and van den Berg and Lindeboom (1998).

Our initial idea was to develop our model taking as a basis the joint probability density function of (T_s, T_a, T_c) conditional on (X, V). However, given the available data, the attrition process could not have been modelled as a duration variable but, instead, it was modelled as a logit specification. Different models can be built depending upon the way we characterize the treatment effect . We choose the simplest representation for the treatment effect by assuming that it is just a constant that multiplies the hazard function of recidivism. Our model resembles the model developed in Abbring and den Berg (2000a). Accordingly, the realization of T_s, say, t_s, affects recidivism behavior, T_c, from t_s onwards, in a deterministic way and only for treatments.

2.5.2 Model

Let $\mathcal{T} = (T_s, T_c) \in \mathbb{R}_+^2$ be a vector of durations. Let $V \in \mathbb{R}_+$ be a random variable representing unobserved heterogeneity[28]. In the absence of treatment effects the model could follow a standard approach: conditional on its unobserved heterogeneity and the exogenous variables, the durations are independent. As usual in the literature, we use the hazard function approach. Adopting a proportional representation for the hazard functions, the econometric model would restrict to specify the following hazard functions[29]:

$$\theta_s(t|X, V) = \lambda_s(t) \cdot \phi_s(X) \cdot V. \tag{2.35}$$

$$\theta_c(t|X, V) = \lambda_c(t) \cdot \phi_c(X) \cdot V. \tag{2.36}$$

Then the conditional survival functions, given X and V, for each of the durations T_s, T_a, T_c is

[28] The gain in using a vector $V = (V_s, V_c)$ is not justified given the purposes of our paper. Hence, for the sake of simplicity, we choose a common random variable V. However, we see the incorporation of $V = (V_s, V_c)$ as an interesting future topic to be explored.

[29] Since there is no treatment effect, the distinction between controls and treatments is meaningless.

$$S_s(t|X,V) = P(T_s \geqslant t|X,V,W) = \exp\left(-V \cdot \Lambda_s(t|X)\right) \tag{2.37}$$

$$S_c(t|X,V) = P(T_c \geqslant t|X,V,W) = \exp\left(-V \cdot \Lambda_c(t|X)\right)$$

where $\Lambda_m(t_m|X)$, $m \in \{s,c\}$, are the integrated hazards:

$$\Lambda_s(t|X) = \phi_s(X) \cdot \int_0^t \lambda_s(\tau)d\tau \tag{2.38}$$

$$\Lambda_c(t|X) = \phi_c(X) \cdot \int_0^t \lambda_c(\tau)d\tau$$

Hence, the joint conditional survival function is:

$$S(t_s,,t_c|X,V) = P\left[T_s \geq t_s, T_c \geq t_c|X,V\right] \tag{2.39}$$

$$= \exp\left(-V \cdot \left(\Lambda_s(t_s|X) + \Lambda_c(t_c|X)\right)\right) \tag{2.40}$$

Finally, in order 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:

$$\begin{aligned} S(t_s,t_c|X) &= \int_0^\infty \exp\left(-v \cdot \left(\Lambda_s(t_s|X) + \Lambda_c(t_c|X)\right)\right) dG(v) \\ &= \mathcal{L}\left(\Lambda_s(t_s|X) + \Lambda_c(t_c|X)\right) \end{aligned}$$

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

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

The key issue now is how to incorporate treatment effects in this framework. A minimal set of conditions necessary to make the effect of treatment a meaningful concept is that:

1. The individual has to be selected in the treatment group;
2. The job search should have ended before the first arrest: $T_s < T_c$.

The dummy W represents group participation and $W = 1$ if the individual belongs to the treatment group. Hence, the conditions for receiving treatment are $W = 1$ and $T_c > T_s$. The problem is now that due to the latter condition it is impossible to build in the effect of treatment directly in the joint survival function[30] without sacrificing the conditional independence of T_s, T_c given X and V. However, note that without assuming conditional independence we can still factorize out the joint density of T_s, T_c conditional on X, V, and W, as a product of conditional densities, say:

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

[30] See Equation (2.39).

Consequently, the corresponding joint survival function can be written as

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

Therefore, in modeling the joint survival function of T_s, 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

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

where $I()$ is the indicators function. If $W = 0$ this specification corresponds to the previous one in (2.36), but for $W = 1$ the effect of the treatment on recidivism is now incorporated:

$$\theta_c(t_c|t_s, t_a, X, V, W = 1) = \phi_c^*(X) \cdot \lambda_c(t_c) \cdot V \; if \; t_s < t_c$$
$$\theta_c(t_c|t_s, t_a, X, V, W = 1) = \phi_c(X) \cdot \lambda_c(t_c) \cdot V \; if \; not,$$

where $\phi_c(X)$ is the same as in (2.36), and $\phi_c^*(X)$ is the systematic hazard during treatment. The corresponding conditional survival functions 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

$$\begin{aligned} &\Lambda_c(t_c|t_s, X, W) \qquad (2.41) \\ = \; & \phi_c(X) \int_0^{t_c} \lambda_c(\tau)d\tau + W. \left(\phi_c^*(X) - \phi_c(X)\right) \int_0^{t_c} I(\tau > t_s)\lambda_c(\tau)d\tau \\ = \; & \phi_c(X) \int_0^{t_c} \lambda_c(\tau)d\tau + W. \left(\phi_c^*(X) - \phi_c(X)\right) I(t_c > t_s) \int_{t_s}^{t_c} \lambda_c(\tau)d\tau \end{aligned}$$

is the corresponding integrated hazard.

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

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

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

$$\begin{aligned} & S(t_s, t_c|X, V, W) \qquad (2.42) \\ = \;& P\left[T_c \geq t_c, T_s \geq t_s|X, V, W\right] \\ = \;& \int_{t_s}^{\infty} S_c(t_c|\tau, X, V, W) f_s\left(\tau|X, V, W\right) d\tau \\ = \;& \int_{t_s}^{\infty} \exp\left[-V \cdot \Lambda_c(t_c|\tau, X, W)\right] V \exp\left(-V \cdot \Lambda_s(\tau|X)\right) \phi_s(X) \lambda_s(\tau) d\tau \\ = \;& V.\phi_s(X) \int_{t_s}^{\infty} \exp\left[-V \cdot \left(\Lambda_c(t_c|\tau, X, W) + \Lambda_s(\tau|X)\right)\right] \lambda_s(\tau) d\tau \end{aligned}$$

where the last two equalities follows from

$$\begin{aligned} f_s\left(t|X, V, W\right) &= -\frac{\partial}{\partial t} S_s(t|X, V, W) \qquad (2.43) \\ &= -\frac{\partial}{\partial t} \exp\left(-V \cdot \Lambda_s(t|X)\right) \\ &= V \exp\left(-V \cdot \Lambda_s(t|X)\right) \phi_s(X) \cdot \lambda_s(t) \end{aligned}$$

Before we specify the baseline hazards and the unobserved heterogeneity distribution, it is worth mentioning a key difference between our model and the setup appearing in Abbring and den Berg (2000a), say, the existence of a control group in our case. This turns out to be fundamental for helping in the identification of the treatment effect, for making inferences about estimated parameters and for giving a minimal characterization of the type of effect we are trying to estimate. Such characterization seems to be nonexistent in the literature.

2.5.3 The Log-likelihood Function

The baseline hazards are assumed to have a Weibull specification[31]. That was the choice of van den Berg, Lindeboom, and Ridder (1994). For the search (unemployment) duration, the Weibull hazards is flexible enough to capture any pattern of monotonic dependence typical of labor markets. Regarding criminal behavior, a "parabola" type hazards would be more appropriate: generally, after release, the ex-criminal has a period of low criminal activity followed by a high criminal activity period. However, as long as the abscissa of the point of maximum of that parabolic hazards is close enough to the origin, the Weibull hazards is still a reasonable approximation. So, we have the following expressions for baseline hazards:

$$\lambda_j(t) = \lambda_j \cdot t^{\lambda_j - 1} \qquad \text{for } j = \{s, c\} \tag{2.44}$$

Then

$$\begin{aligned} & \Lambda_c(t_c | t_s, X, W = 1) \\ = \quad & \phi_c(X) t_c^{\lambda_c} + \left(\phi_c^*(X) - \phi_c(X)\right) I(t_c > t_s) \left(t_c^{\lambda_c} - t_s^{\lambda_c}\right) \end{aligned} \tag{2.45}$$

and

$$\Lambda_s(t|X) = \phi_s(X) \cdot t^{\lambda_s} \tag{2.46}$$

---

[31]Of course, the estimation can proceed without imposing any functional form on the baseline hazards. However, we believe our sample is not large enough to taken semiparametric estimation seriously.

Hence, it follows from Appendix B.1 that:

$$
\begin{aligned}
& S(t_s, t_c|X, V, W = 1) \qquad (2.47) \\
= \;& I(t_c > t_s)V.\phi_s(X) \exp\left(-V \cdot \phi_c(X) t_c^{\lambda_c}\right) \\
& \times \int_0^{\infty} I(t_s < \tau < t_c) \exp\left[-V \cdot \left(\phi_c^*(X) - \phi_c(X)\right)\left(t_c^{\lambda_c} - \tau^{\lambda_c}\right)\right] \\
& \times \exp\left[-V \cdot \left(\phi_s(X) \cdot \tau^{\lambda_s}\right)\right] \lambda_s \tau^{\lambda_s - 1} d\tau \\
& + I(t_c > t_s) \exp\left[-V \cdot \left(\phi_c(X) t_c^{\lambda_c} + \phi_s(X) \cdot t_c^{\lambda_s}\right)\right] \\
& + I(t_c \leq t_s) \exp\left[-V \cdot \left(\phi_c(X) t_c^{\lambda_c} + \phi_s(X) \cdot {t_s}^{\lambda_s}\right)\right]
\end{aligned}
$$

Where we have the following for controls:

$$
\begin{aligned}
& S(t_s, t_c|X, V, W = 0) \qquad (2.48) \\
= \;& \exp\left[-V \cdot \left(\phi_c(X) \cdot t_c^{\lambda_c} + \phi_s(X) \cdot t_s^{\lambda_s}\right)\right]
\end{aligned}
$$

In order to be able to compute all the probability expressions above, we need also to specify a distribution for the unobserved heterogeneity. For the sake of simplicity again, the distribution of V is assumed to be gamma. This choice has been traditional in the literature, for instances, see Lancaster (1990). This so because of its nice features, specially because its Laplace transform is a closed form expression. If $V \sim Gamma(\alpha, \omega)$ then the Laplace transform of V is:

$$\mathcal{L}(s) = E\left[\exp(-s.V)\right] = \left(1 + s \cdot \omega\right)^{-\alpha}, \tag{2.49}$$

with derivative

$$\mathcal{L}'(s) = -E\left[V \exp(-s.V)\right] = -\alpha\omega \left(1 + s \cdot \omega\right)^{-\alpha-1}. \tag{2.50}$$

Therefore, it follows from (2.47)- (2.50) that:

$$
\begin{aligned}
& S(t_s, t_c|X, W = 1) && (2.51)\\
= \; & I(t_c > t_s)\alpha\omega\phi_s(X)\int_0^{\infty} I(t_s < \tau < t_c) \\
& \times \left[1 + \omega\left(\phi_c(X)t_c^{\lambda_c} + \left(\phi_c^*(X) - \phi_c(X)\right)\left(t_c^{\lambda_c} - \tau^{\lambda_c}\right) + \phi_s(X)\cdot\tau^{\lambda_s}\right)\right]^{-(\alpha+1)} \\
& \times\lambda_s\tau^{\lambda_s - 1}d\tau \\
& + I(t_c > t_s)\left[1 + \omega\left(\phi_c(X)t_c^{\lambda_c} + \phi_s(X)\cdot t_c^{\lambda_s}\right)\right]^{-\alpha} \\
& + I(t_c \leq t_s)\left[1 + \omega\left(\phi_c(X)t_c^{\lambda_c} + \phi_s(X)\cdot {t_s}^{\lambda_s}\right)\right]^{-\alpha}
\end{aligned}
$$

and

$$
S(t_s, t_c|X, W = 0) = \left[1 + \omega\left(\phi_c(X)\cdot t_c^{\lambda_c} + \phi_s(X)\cdot t_s^{\lambda_s}\right)\right]^{-\alpha} \tag{2.52}
$$

The integral in (2.51) can be further “simplified” as:

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

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 to have the integral well-defined, it must be the case that:

$$a > 0,\ b \geq 0,\ p \geq 0,\ q \geq p,\ r \geq 0,\ \text{and}\ c > -\frac{b+p}{q^r - p^r}. \tag{2.53}$$

Before we build the likelihood function, it is important to discuss how we are going to deal with some limitations of the sample. A first problem is posed by the group of no shows. Given the features of our program, no shows exist in both the control and treatment group. There are two basic approaches to deal with this: either one sticks to the duration approach and model no shows as independent right censoring with censoring time close but different of zero or apply a discrete choice model. However, it is a well established fact in the literature of microeconometrics[32] that those two choices are different.

A second issue is related to the attrition process. The data set reveals only if somebody drops out before or after placement. Since, for the sake of simplicity, we decided not to consider attrition after placement, there is only the need to figure out the attrition date before placement. Even though dropping out does not imply no observation of the duration of recidivism, we are not going to make use of this information available in our sample. Hence, we need to obtain the following expression:

$$P_j(DO = 1|x) \equiv P_j(T_a < T_s|x) \tag{2.54}$$

The last probability is equivalent:

[32] See, for instances Pudney (1989).

$$P_j(DO = 1|x) = \int_0^{\infty} P_j(T_a \in [0,\tau)|T_s = \tau, x) \cdot f_s^j(\tau) \cdot d\tau \quad (2.55)$$

Where $f_s^j(t)$ is the probability density function of T_s for controls and treatments. The expressions appearing inside the integral can be readily available from the survival function. Nonetheless, a closed expression for the integral itself is highly unlikely even for very simple hazard functions and unobserved heterogeneity distributions. Thus, simulation or approximation appears to be the only sensible ways to deal with that. We described in Appendix B.3 the numerical approach for the calculation of the probability of interest. To conclude, we think that the motivations for dropping out are likely different for each group, thus the parameters in the probability expression are not necessarily the same for controls and treatments.

In view of the discussion about "no shows" and attrition, we decided to take a very pragmatic approach. Instead of modelling these two issues separately, we assume that this model applies conditionally on the absence of attrition, where attrition now also include "no show". Thus, let $N = 1$ if the individual quits before the treatment is completed, or before finding a job, or do not show up at all, and $N = 0$ otherwise. Then specify

$$P[N = 1|X, W = w] = \frac{1}{1 + \exp\left(-\gamma_w' X\right)}, \; w = 0, 1, \quad (2.56)$$

so that

$$P[N = 0|X, W = w] = \frac{\exp\left(-\gamma_w' X\right)}{1 + \exp\left(-\gamma_w' X\right)} = \frac{1}{1 + \exp\left(\gamma_w' X\right)}, \;\; w = 0, 1, \tag{2.57}$$

The parameters γ_w may be different for $w = 0, 1$. The duration model is then conditional on $N = 0$.

A final question is how to deal with a sample where the survival times are only known to belong to intervals. In another words, we need to deal with grouped observations. Fortunately, there are only 6 different realizations of interest. We describe below the way to compute the probabilities for each of these possible realizations:

1. **Arrest and placement**

   This includes three possibilities:

   (a) Arrest before placement

$$P(T_s \in [t_1, t_2), T_c \in [t_3, t_4)|x) \qquad \text{where } t_1 > t_4 \tag{2.58}$$

   (b) Arrest after placement

$$P(T_s \in [t_1, t_2), T_c \in [t_3, t_4)|x) \qquad \text{where } t_3 > t_2 \tag{2.59}$$

(c) Unordered arrest

This happens when there is an intersection between the interval of realization.

The probability we are trying to calculate is the following:

$$P(T_s \in [t_1, t_2), T_c \in [t_1, t_3)|x) \tag{2.60}$$

2. **Censored arrest**

$$P(T_s \in [t_1, t_2), T_c \in [\bar{t}, \infty)|x) \qquad \text{where } \bar{t} > t_2 \tag{2.61}$$

3. **Censored Search**

$$P(T_s \in [\bar{t}, \infty), T_c \in [t_3, t_4)|x) \tag{2.62}$$

4. **Others**

This category comprises any individual realization that does not belong to any of the five categories above. The key difference between **Others** and the categories above is that the interval of T_s and T_c does not belong to any of the twelve possible disjoint realizations defined in Appendix B.2.

To compute the probabilities outlined above one needs to apply the decomposition that represents survival functions in terms of linear combinations of distribution functions. $P(T_s \in [a,b), T_c \in [c,d))$ can be represented by means of the cumulative distribution function , $\mathbf{F}(m,n) \equiv Prob(T_s < m, T_c < n)$, as:

$$P(T_s \in [a,b), T_c \in [c,d)) = \mathbf{F}(b,d) - \mathbf{F}(b,c) - \mathbf{F}(a,d) + \mathbf{F}(a,c) \tag{2.63}$$

Such decomposition makes possible the calculation of the probability of unordered arrest. Remember that this is the case when it is not possible to order the realizations of recidivism and placement time. However, by way of Equation (2.63) this is just an apparent problem because there is no need to know which event happened first. Another interesting decomposition of $P(T_s \in [a,b), T_c \in [c,d))$ in terms of survival function was suggested by Prof. Herman Bierens, from PennState University. It decomposes the probability into survival functions that have intervals that are either nonoverlapping or common. See Appendix B.4 for details.

To accommodate all possible realizations, we develop the following notation. 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 $i = 1, ..., k$,

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

say, where $W_j = 0$ if individual j belongs to the control group, and $W_j = 1$ to the treatment group. Moreover, the probability of an individual belonging to the category **Others** is[33]:

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

say.

Next, let $N_j = 1$ if individual j does not show up, or quit, with probability

$$P\left[N_j = 1|X_j, W_j = i\right] = q_j(\gamma_i),\ i = 0, 1$$

say. It is assumed that

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

[33]Note that we are not using any information about T_s and T_c for individuals in this category, even though we do have it.

Then the log-likelihood takes the form:

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

say, where n is the sample size.

Given the structure of the likelihood, estimation will be performed separately. Parameters estimates from both $\log \mathcal{L}_0(\gamma_0)$ and $\log \mathcal{L}_1(\gamma_1)$ can be obtained from a logit model. Since individuals are randomly selected to be either controls or treatments, the estimation procedure should follow these steps:

1. Split the sample between controls and treatments. Use the indicator W to do that;
2. Transform the variables accordingly and run a logit. The dependent variable will be N and the set of regressors X and W;
3. Obtain $\widehat{\gamma_0}$ and $\widehat{\gamma_1}$.
4. Finally, we use maximum likelihood estimation to estimate the parameters of $\log \mathcal{L}_2(\beta)$.

The result of the estimations as well as inferences appears in the next section. It is worth saying that all econometric work (data manipulation, estimation and inference) was conducted by means of an econometric package **EasyReg**, written by Prof. Herman Bierens from PennState University. This freeware program can be downloaded from the following website: **http://econ.la.psu.edu/~hbierens/EASYREG.HTM** .

2.5.4 Estimation and Inference

Since we are going to estimate both vector of regressors γ_0 and γ_1 simultaneously, we adjust the set of regressors X by way of the group dummy W. Hence, the actually estimated logit specification is:

$$P[N = 1|X, W = \{0; 1\}] = \frac{1}{1 + \exp\big(- (1 - W) \cdot X'\gamma_0 - W \cdot X'\gamma_1 \big)}. \tag{2.64}$$

The results for the logit regression appear in table 2.3. Note that given the logit specification, the second set of estimated parameters corresponds to the difference between the estimated parameters of treatments and controls. The majority of the estimated regressors are not significant at the 10% level[34]. A likely explanation for that relies on the fact that we have not distinguished between "no-shows" and attrition. It is possible that "no-shows" and those who have experienced attrition require different set of regressors and/or different models. However, as explained before, the inclusion of attrition would mean an additional duration variable which would complicate an already complex model. Hence, the logit specification and the aggregation of "no-shows" and attrition seems to be a reasonable choice in face of our available trade-off. As a matter of fact, it should be clear that our focus is on the estimation of the main model.

For controls, only variables sex, chicago and sandiego are significant. A man has a higher probability of attrition than a woman does. Belonging to the program located in Chicago, as well as in San Diego, raises the probability of attrition. For treatments, the results are very similar. Since our principal interest lies on the parameters of the survival functions, we do not discuss this last results. An apparent pattern is the influence of the program's structure, as captured by the program's site dummies. So, the program itself might have greater influence on the attrition process than individual characteristics. The intercept is also significant. Finally, we test if there is any difference between

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

the parameters of controls and treatments. We tested for equality of the estimated parameters that are significant for both controls and treatments:

$$H_0 : \gamma_{(1,.)} = \gamma_{(0,.)} \qquad H_1 : \gamma_{(1,.)} \neq \gamma_{(0,.)}. \tag{2.65}$$

In the present case, it is sufficient to test if the second set of estimated parameters (treatments - controls) is significantly different from zero. All estimated parameters, except chicago, are not significantly different from zero. It appears that controls and treatments have similar behavior regarding attrition from the program. This is not entirely surprising given the randomization occurred during group's assignment. We turn to the estimation of the main model.

Results appear in table 2.4. The first subscript denotes either search, s, or crime, c. They were obtained by maximizing the likelihood function by the simplex method of Nelder and Meade[35] which does not rely on derivatives. As a general assessment, judging by the t-values, the model performs reasonably well. All coefficients with the exception of $\beta_{(s,firstarest)}$, $\beta_{(s,drug)}$, $\beta_{(s,race)}$, and $\beta_{(c,firstarest)}$, $\beta_{(c,educ)}$, $\beta_{(c,chicago)}$ and $\beta_{(c,sandiego)}$ are significant. Before going into details, it is worth to clarify the effect of the regressors as measured by $\beta_{(.,.)}$. Given the proportional specification of the hazards function and the exponential form of the systematic hazards function, the effect of a regressor on the dependent variable is:

- $\beta > 0$ and $X > 0$: The whole hazards function is inflated if X goes up. So, expected time of failure is shifted to the left. In the case of search duration, a

[35] See, Press, Flannery, Teukolsky, and Vetterling (1989).

positive β means that the average time of search (unemployment) is lower the higher the value of X. For the crime (recidivism) duration this means that the expected recidivism time shifts left, therefore anticipating the criminal activity[36];

- $\beta < 0$ and $X > 0$: The hazards function is deflated if X goes up and the effects can be deduced accordingly;

- $\beta = 0$: There is no effect.

We begin by discussing the first set of estimates. As it will be clear from the estimated parameters for the search duration, some results appear to contradict well established facts in the literature of empirical search models. However, given the specific nature of our model, say, job search program for 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, therefore it increases the expected search time the older the ex-inmate is. At first, this result does not contradict the extensive evidence that older people have a tougher time trying to find a jog than young people do. Neither it is difficult to justify it in terms of a precautionary behavior from the demand side: employers might well be aware that old criminals might have a tough time trying to go "straight". Hence, older ex-criminals trying to abandon their criminal careers

[36]Of course, if $X < 0$ the effects are inverted for both durations.

will face a rather adverse environment. Nonetheless, a value of $\beta_{(s,age)} = -0.07566$, although highly significant, does not represent an important effect when compared to others regressors.

Males appears to have search time greater than the search time of women. This is the first result that contradicts empirical findings in search models. Indeed, the sex effect is strong, $\beta_{(s,educ)} = -1.147$. However, males have a much higher criminal behavior and, consequently, presents a higher potential threat inside a job. The positive effect of education, $\beta_{(s,educ)} = 0.348$, means that more educated ex-inmates will find jobs faster than less educated. In order to assess the impact of local labor market conditions and program heterogeneity, the coefficients of two dummies, one for Chicago and the other for San Diego, are estimated. As their estimated parameter values and t-statistics indicate, they are both strongly negative and significant. From the values of $\beta_{(s,chicago)} = -2.021$ and $\beta_{(s,sandiego)} = -1.071$, we can conclude that the local conditions of Chicago and San Diego[37] do not help ex-inmates find jobs faster, while in Boston the effect is opposite. 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 keep searching the higher the probability of finding a job. This is exactly the opposite of a lot of evidence found in studies of search in the labor market. For instance, see Devine and Kiefer (1991). A very plausible explanation for the positive and highly significant λ_s is that very soon the ex-inmate will know how difficult is to find a reasonable job, given the strong stigma

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

prevalent in the society. Hence, sooner or later he/she will have no option except to accept the low-paying job. We turn now to the analysis of the recidivism.

The impact of age in the expected recidivism time is negative, $\beta_{(c,age)} = -0.117$. 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. Past drug use shows a surprising negative value. So, past drug users are less likely to commit a crime earlier than non-drug users. The estimated parameter of the variable race is positive, in accordance of a huge amount of evidence. Hence, non-whites tend to recidivate earlier than whites. The estimated effect of sex, $\beta_{(c,sex)} = -1.670$, is among the greatest and appears to contradicts the literature on criminal recidivism: females will commit a crime earlier than males. The inclusion of city dummies did not prove to be a strategy worth pursuing, as demonstrated by the non-significant t-values of Chicago and San Diego. Finally, a $\lambda_c = 1.31812$ shows that recidivism process presents positive dependence, or, the longer you stay without committing a crime the higher the probability of committing it.

The parameter α measures the impact of the unobservable heterogeneity on the conditional hazards. Since $\omega = 1$, α is a consistent estimator of $\mathbf{E}(V)$, the mean of the gamma unobserved heterogeneity random variable. Thus, a nice interpretation for α, as suggested by equations (2.35) and (2.36), is that it modifies both hazard functions by either inflating or deflating them. The interesting test should be then:

$$H_0 : \alpha = 1 \qquad H_1 : \alpha \neq 1. \tag{2.66}$$

A $t-test$ rejects the null[38]. The estimated effect of the unobserved heterogeneity is not significantly different from one. Hence, at least in our set up, unobserved heterogeneity is not an important issue.

The parameter δ is the key parameter on our econometric model. As explained before, it measures the effect of the ESEO program on the recidivism behavior of its participants. It works by either inflating or deflating the portion of the hazards function for recidivism conditional on placement. Thus, its interpretation is straightforward:

- If $\delta > 1$, the program has a negative impact on recidivism, as it inflates the hazards for recidivism, therefore anticipating the expected time of committing a crime;
- 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 postponing the expected time of committing a crime.

Again, the right test should be analogous to the test performed for α. Since the t-statistics is -2.306, the null hypothesis is rejected. Hence, the program is **effective** ($\delta = 0.631$) as it attenuates the criminal activity of those who received its services. This last result stands in contrast with the original study of the ESEO program, as shown in Mylkman, Timrots, Peyser, Toborg, Yezer, Carpenter, and Landson (1985).

[38] The value of the t-statistic is $\frac{1.21336-1}{0.3945} = 0.540$

2.6 Conclusion

We believe that the econometric study of the effects of the ESEO program on the recidivism behavior of ex-criminals represents a modest, but important contribution to the literature on econometric evaluation of social programs. By modelling the ESEO program as a mixed multivariate proportional hazards model[39], where treatment is conditional on placement, we have put together 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[40]. 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 80'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 2.4, 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 constraint on what can be identified from the model and makes our results less

[39] See, Abbring and den Berg (2000b) for this nomenclature.

[40] Of course, the importance of the theoretical paper of Abbring and den Berg (2000a) should be made explicit. This is so, even though we go well beyond them by expanding their framework, by estimating parameters and, hopefully, by making clear the structure of the dependence between the search and crime durations.

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, Lalonde, and Smith (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 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.

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. 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.

This lead us to suggest an urgent new avenue to explore: build models that assume impact heterogeneity. From the perspective of our model this means to specify the following:

$$\delta(X) = J(X) \qquad \text{where } J(X) \geq 0 \quad \text{for all } X \in \mathbb{R}^n. \tag{2.67}$$

Thus, the effect of the program is conditional on a set of regressors representing individual-specific, program-specific and local variables. Undoubtedly, this would give us a much more accurate picture of the program. As a matter of fact, an easy choice would be $\delta(X) = \exp[X'\beta]$. However, identification of the model becomes a problem!

The *Handbook of Labor Economics* chapter on econometric evaluation of active labor market programs shows how well-developed and active remains this topic of research. Also, the *Handbook of Econometrics* chapter on duration analysis reaches similar conclusion. Such intersection unequivocally opens an exciting new area of research where the time dimension of programs are studied with further detail. This new approach will definitely result on the advancement of our understanding of how social programs work.

Table 2.3. Logit Parameters Estimate:

Group	Parameters	ML estimate	t-values
controls	age	0.022	1.11
	first arrest	-0.040	-1.36
	drug	0.120	0.48
	race	-0.275	-1.08
	sex	1.006	2.85
	education	-0.176	-1.11
	chicago	1.884	6.99
	sandiego	1.205	3.74
	intercept	-0.741	-0.95
treatments - controls	age	0.011	0.87
	first arrest	0.032	1.76
	drug	0.062	0.34
	race	0.222	1.18
	sex	0.036	0.13
	education	-0.194	-1.64
	chicago	0.548	2.50
	sandiego	0.373	1.67
	intercept	-0.761	-0.81
	Log-Likelihood		-647.686
	n		1074

Table 2.4. 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

Appendix A

A.1 The Model

Let for $j = 1, 2$,

$$P[T_j > t|V] = \exp\left[-\int_0^t \theta(\tau, \beta_j)d\tau.V\right],$$

where $\theta(\tau, \beta_j)$ is the hazard function and the β_j's are parameter vectors of equal dimension. Conditional on $V > 0$, T_1 and T_2 are independent. The random variable V has Laplace transform

$$\Psi(\lambda) = E[\exp(-\lambda V)],\ \lambda > 0,$$

where α is a parameter vector. The T_j's are only observed if $T_j \leq \overline{T}$.

Let

$$C = I\left[T_1 > \overline{T}\right].I\left[T_2 > \overline{T}\right] = I\left[\min(T_1, T_2) > \overline{T}\right],$$

where $I(\cdot)$ is the indicator function. Then

$$\begin{aligned} P[C &= 1] = E\left(P[T_1 > \overline{T}|V].P[T_2 > \overline{T}|V]\right) \\ &= \Psi\left(\int_0^{\overline{T}} \theta(\tau, \beta_1)d\tau + \int_0^{\overline{T}} \theta(\tau, \beta_2)d\tau \middle| \alpha\right). \end{aligned}$$

This is the probability of censoring.

Given that $C = 0$, we observe $T = \min(T_1, T_2)$, and $F = I\left(T = T_2\right)$. I shall now derive the conditional density of T, given F and $C = 0$.

A.1.1 The case $F = 0$

We have

$$P\left[T > t | F = 0, C = 0\right].P[F = 0, C = 0]$$

$$\begin{aligned}
&= P\left[T_1 > t, T_2 > t, F = 0, C = 0\right] \\
&= P\left[T_1 > t, T_2 > t, \min(T_1, T_2) = T_1, \min(T_1, T_2) \leq \overline{T}\right] \\
&= P\left[T_1 > t, T_2 > t, T_2 > T_1, T_1 \leq \overline{T}\right] \\
&= P\left[t < T_1 \leq \overline{T}, T_2 > \max(t, T_1)\right] \\
&= E\left(E\left[I\left(t < T_1 \leq \overline{T}\right) I\left(T_2 > \max(t, T_1)\right) \middle| V, T_1\right]\right) \\
&= E\left[I\left(t < T_1 \leq \overline{T}\right) P\left(T_2 > \max(t, T_1) \middle| V, T_1\right)\right] \\
&= E\left[\int_t^{\overline{T}} \exp\left[-\int_0^{\max(t,\tau)} \theta(\tau, \beta_2) d\tau.V\right] dP\left[T_1 \leq \tau | V\right]\right] \\
&= -E\left[\int_t^{\overline{T}} \exp\left[-\int_0^{\tau} \theta(\varsigma, \beta_2) d\varsigma.V\right] \frac{dP\left[T_1 > \tau | V\right]}{d\tau} d\tau\right]
\end{aligned}$$

Substituting

$$\begin{aligned}
\frac{dP\left[T_1 > \tau | V\right]}{d\tau} &= \frac{d\exp\left[-\int_0^{\tau} \theta(\varsigma, \beta_1) d\varsigma.V\right]}{d\tau} \\
&= -V.\theta(\tau, \beta_1).\exp\left[-\int_0^{\tau} \theta(\varsigma, \beta_1) d\varsigma.V\right]
\end{aligned}$$

it follows that

$$
\begin{aligned}
& P\left[T > t|F = 0, C = 0\right] \\
= \quad & \frac{E\left[\int_t^{\overline{T}} V.\theta(\tau,\beta_1).\exp\left[-\int_0^{\tau}\theta(\varsigma,\beta_1)d\varsigma.V - \int_0^{\tau}\theta(\varsigma,\beta_2)d\varsigma.V\right]d\tau\right]}{P[F = 0, C = 0]},
\end{aligned}
$$

hence

$$
\begin{aligned}
f(t|F &= 0, C = 0)P[F = 0, C = 0] \qquad \text{(A.1)} \\
&= -\frac{dP\left[T > t|F = 0, C = 0\right]}{dt} \times P[F = 0, C = 0] \\
&= \theta(t,\beta_1).E\left(V.\exp\left[-\int_0^t \theta(\varsigma,\beta_1)d\varsigma.V - \int_0^t \theta(\varsigma,\beta_2)d\varsigma.V\right]\right) \\
&= -\theta(t,\beta_1)\Psi'\left(\int_0^t \theta(\varsigma,\beta_1)d\varsigma. + \int_0^t \theta(\varsigma,\beta_2)d\varsigma \middle| \alpha\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi'(\lambda|\alpha) &= \partial\Psi(\lambda|\alpha)/\partial\lambda \\
&= \frac{\partial E\left[\exp(-\lambda V)\right].}{\partial\lambda} \\
&= E\left[\frac{\partial \exp(-\lambda V)}{\partial\lambda}\right] \\
&= -E\left[V\exp(-\lambda V)\right].
\end{aligned}
$$

Note that there is no need to derive $P[F = 0, C = 0]$, as (A.1) is all we need in the case $F = 0$. See the section on the log-likelihood below. This is fortunate, because $P[F = 0, C = 0]$ is a nasty integral:

$$
\begin{aligned}
P[F &= 0, C = 0] \qquad\qquad (A.2)\\
&= P\left[\min(T_1, T_2) = T_1, \min(T_1, T_2) \leq \overline{T}\right]\\
&= P\left[T_2 > T_1, T_1 \leq \overline{T}\right]\\
&= E\left(I(T_1 \leq \overline{T}).I(T_2 > T_1)\right)\\
&= E\left(I(T_1 \leq \overline{T}).I(T_2 > T_1)\middle|\, T_1, V\right)\\
&= E\left(I(T_1 \leq \overline{T}).P(T_2 > T_1|\, V)\right)\\
&= E\left(\int_0^{\overline{T}} P(T_2 > t|\, V) dP(T_1 \leq t|\, V)\right)\\
&= -E\left(\int_0^{\overline{T}} P(T_2 > t|\, V)\frac{dP(T_1 > t|\, V)}{dt}dt\right)\\
&= \int_0^{\overline{T}} \theta(t, \beta_1) E\left(V \exp\left[-\int_0^t \theta(\tau, \beta_1) d\tau.V - \int_0^t \theta(\tau, \beta_2) d\tau.V\right]\right) dt\\
&= \int_0^{\overline{T}} \theta(t, \beta_1)\Psi'\left(\int_0^t \theta(\varsigma, \beta_1) d\varsigma. + \int_0^t \theta(\varsigma, \beta_2) d\varsigma\middle|\, \alpha\right) dt
\end{aligned}
$$

A.1.2 The case $F = 1$

The case $F = 1$ follows straightforwardly from the case $F = 0$, simply by swapping β_1 and β_2. Thus.

$$
\begin{aligned}
f(t|F &= 1, C = 0)P[F = 1, C = 0] \qquad\qquad (A.3)\\
&= -\theta(t, \beta_2)\Psi'\left(\int_0^t \theta(\varsigma, \beta_1) d\varsigma. + \int_0^t \theta(\varsigma, \beta_2) d\varsigma\middle|\, \alpha\right)
\end{aligned}
$$

and

$$\begin{aligned} P[F &= 1, C = 0] && \text{(A.4)} \\ &= \int_0^{\overline{T}} \theta(t,\beta_2)\Psi'\left(\int_0^t \theta(\varsigma,\beta_1)d\varsigma. + \int_0^t \theta(\varsigma,\beta_2)d\varsigma \middle| \alpha \right) dt. \end{aligned}$$

Again, (A.3) is all we need in the case $F = 1$; we don't need the probability (A.4).

A.2 Parameter Identification

Let us look at the first-order conditions. First, let

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}(\beta_1,\beta_2,\gamma,\delta)}{\partial\gamma} \\ &= \frac{1}{\gamma}\sum_{j=1}^{n}\left(1-C_j\right) - \sum_{j=1}^{n}\ln\left(1+\delta\left(\int_0^{\overline{T}_j}\theta(\tau,\beta_1)d\tau + \int_0^{\overline{T_j}}\theta(\tau,\beta_2)d\tau\right)\right). \end{aligned}$$

Then

$$\gamma = \frac{\sum_{j=1}^{n}\left(1-C_j\right)}{\sum_{j=1}^{n}\ln\left(1+\delta\left(\int_0^{\overline{T}_j}\theta(\tau,\beta_1)d\tau + \int_0^{\overline{T_j}}\theta(\tau,\beta_2)d\tau\right)\right)}. \tag{A.5}$$

Next, let

$$
\begin{aligned}
0 &= \frac{\partial \mathcal{L}(\beta_1, \beta_2, \gamma, \delta)}{\partial \delta} \qquad \text{(A.6)} \\
&= \frac{1}{\delta} \sum_{j=1}^{n} \left(1 - C_j\right) \\
&\quad + \sum_{j=1}^{n} C_j \left(\frac{\left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)}{1 + \delta \left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)} \right) \\
&\quad - (1+\gamma) \sum_{j=1}^{n} \left(\frac{\left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)}{1 + \delta \left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)} \right)
\end{aligned}
$$

Substituting (A.5) in (A.6) yields

$$
\begin{aligned}
0 &= \frac{1}{\delta} \sum_{j=1}^{n} \left(1 - C_j\right) \\
&\quad + \sum_{j=1}^{n} (C_j - 1) \left(\frac{\left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)}{1 + \delta \left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)} \right) \\
&\quad - \frac{\sum_{j=1}^{n} \left(1 - C_j\right)}{\sum_{j=1}^{n} \ln \left(1 + \delta \left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)\right)} \\
&\quad \times \sum_{j=1}^{n} \left(\frac{\left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)}{1 + \delta \left(\int_0^{\overline{T}_j} \theta(\tau, \beta_1) d\tau + \int_0^{\overline{T_j}} \theta(\tau, \beta_2) d\tau\right)} \right),
\end{aligned}
$$

which can be solved for δ, given β_1 and β_2, and substituting this solution in (A.5), γ is determined, given β_1 and β_2. Thus, γ and δ are identified, so that there is no need to normalize δ to 1!

Appendix B

B.1 Joint Survival Function

It follows from equations (2.42), (2.45) and (2.46) that:

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

Finally,

$$
\begin{aligned}
& S(t_s, t_c|X, V, W = 1) && \text{(B.1)} \\
= \quad & I(t_c > t_s)V.\phi_s(X) \exp\left(-V \cdot \phi_c(X) t_c^{\lambda_c}\right) \\
& \times \int_0^{\infty} I(t_s < \tau < t_c) \exp\left[-V \cdot \left(\phi_c^*(X) - \phi_c(X)\right)\left(t_c^{\lambda_c} - \tau^{\lambda_c}\right)\right] \\
& \times \exp\left[-V \cdot \left(\phi_s(X) \cdot \tau^{\lambda_s}\right)\right] \lambda_s \tau^{\lambda_s - 1} d\tau \\
& + I(t_c > t_s) \exp\left[-V \cdot \left(\phi_c(X) t_c^{\lambda_c} + \phi_s(X) \cdot t_c^{\lambda_s}\right)\right] \\
& + I(t_c \leq t_s) \exp\left[-V \cdot \left(\phi_c(X) t_c^{\lambda_c} + \phi_s(X) \cdot {t_s}^{\lambda_s}\right)\right]
\end{aligned}
$$

Where we have the following for controls:

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

B.2 Duration Variables

The variable T_s measures the search time, or the time between placement date and start up date (the later between release date and program starting date). See Table B.1.

Table B.1. Length of Search Time - T_s

T_s	**Interval**	**Days**
0	(0,1)	< 1
1	[1, 15)	15
2	[15, 30)	15
3	[31, 60)	30
4	[61, 120)	60
5	[121, 180)	60
6	[181, 240)	60
7	[241, 300)	60
8	[301, 360)	60
9	[361, 480)	120
10	[481, ∞)	$\geqslant 481$
88	not applicable	
99	no information	

The variable T_c measures the recidivism time, or the time necessary to be arrested since the start up date. It is a derived variable and is shown in Table B.2.

For the sake of simplicity, we transform the original intervals of T_c and T_s into a set of non-overlapping and adjacent intervals. Thus, both durations are measured by the following set of intervals[1]:

Hence, there are 16 possible realizations of $T_s \times T_c$ from which 12 have non-overlapping endpoints, say, $T_s \times T_c \in \{(1,2),(1,3),\cdots,(4,3)\}$, where $(i,j) : i = 1,2,3$ and 4 and $j = 1,2,3$ and 4 are interval numbers.

B.3 Probability of Drop Out for Weibull Specification

B.3.1 Analytical Expression

Let the hazard functions for search time and attrition have a Weibull form:

[1]If the individual belongs to the group of **Others**, T_c is coded 5.

Table B.2. Recidivism Time - T_c

T_c	**Interval**	**Days**
1	(0,30]	30
3	(30, 90]	60
6	(90, 180]	90
12	[180, 360]	180
18	(360, 540]	180
24	(540, 720]	180
30	(720, 900]	180
36	(900, 1080]	180

$$\theta_s(t|x,v) = \lambda_s \cdot t^{\lambda_s - 1} \cdot e^{\beta_s' x} \cdot v \tag{B.3}$$

$$\theta_a(t|x,v) = \lambda_a \cdot t^{\lambda_a - 1} \cdot e^{\beta_a' x} \cdot v \tag{B.4}$$

It is easy to see that the probability of drop out is:

$$P_j(DO = 1|x) = 1 - \int_0^\infty \alpha\omega e^{\beta_s' x} \lambda_s t^{\lambda_s - 1} \cdot \Big(1 + \omega \cdot \Big\{ e^{\beta_s' x} \cdot t^{\lambda_s} + e^{\beta_a' x} \cdot t^{\lambda_a} \Big\}\Big)^{-\alpha - 1} \cdot dt \tag{B.5}$$

Table B.3. Common Intervals

Number	**Interval**	**Days**
1	(0,30]	30
2	(30, 180]	150
3	(180, 360]	180
4	(360, ∞)	-

B.3.2 Numerical Approximation

The integral in Equation (B.5) takes the form

$$\int_0^\infty \frac{\alpha p \lambda t^{\lambda-1}}{\left(1 + pt^\lambda + qt^\mu\right)^{\alpha+1}} dt \tag{B.6}$$

where

$$\begin{aligned} p &= \omega.\exp(\beta_s' x) \\ q &= \omega.\exp(\beta_a' x) \\ \lambda &= \lambda_s \\ \mu &= \lambda_a \end{aligned}$$

First note that

$$
\begin{aligned}
& \int_0^\infty \frac{\alpha p \lambda t^{\lambda-1}}{\left(1+pt^\lambda+qt^\mu\right)^{\alpha+1}} dt + \int_0^\infty \frac{\alpha q \mu t^{\mu-1}}{\left(1+pt^\lambda+qt^\mu\right)^{\alpha+1}} dt \qquad \text{(B.7)} \\
= & \int_0^\infty \frac{\alpha p \lambda t^{\lambda-1} + \alpha q \mu t^{\mu-1}}{\left(1+pt^\lambda+qt^\mu\right)^{\alpha+1}} dt = -\int_0^\infty d\left(1+pt^\lambda+qt^\mu\right)^{-\alpha} \\
= & \ 1
\end{aligned}
$$

Second, if $\lambda = \mu$ then it follows from (B.7) that (B.6) becomes

$$
\begin{aligned}
& \int_0^\infty \frac{\alpha p \lambda t^{\lambda-1}}{\left(1+pt^\lambda+qt^\lambda\right)^{\alpha+1}} dt \\
= & \ \frac{p}{p+q} \int_0^\infty \frac{\alpha(p+q)\lambda t^{\lambda-1}}{\left(1+(p+q)t^\lambda\right)^{\alpha+1}} dt = \frac{p}{p+q}.
\end{aligned}
$$

Next assume $\lambda \neq \mu$. The integral (B.6) can be symplified to

$$\begin{aligned}
& \int_0^\infty \frac{\alpha p \lambda t^{\lambda-1}}{\left(1+pt^\lambda+qt^\mu\right)^{\alpha+1}} dt \\
= & \int_0^\infty \frac{\alpha}{\left(1+pt^\lambda+qt^\mu\right)^{\alpha+1}} dpt^\lambda \\
= & \int_0^\infty \frac{\alpha}{\left(1+\xi+r.\xi^{\mu/\lambda}\right)^{\alpha+1}} d\xi \qquad \text{(B.8)}
\end{aligned}$$

where

$$\begin{aligned}
pt^\lambda &= \xi \\
t &= (\xi/p)^{1/\lambda} \\
qt^\mu &= q\,(\xi/p)^{\mu/\lambda} = q.p^{-\mu/\lambda}\xi^{\mu/\lambda} \\
r &= q.p^{-\mu/\lambda}
\end{aligned}$$

say. Moreover, the integral (B.8) can be written as

$$
\begin{aligned}
& \int_0^\infty \frac{\alpha}{\left(1+\xi+r.\xi^{\mu/\lambda}\right)^{\alpha+1}} d\xi \\
= & \int_0^\infty \frac{-\alpha\exp(\xi)}{\left(1+\xi+r.\xi^{\mu/\lambda}\right)^{\alpha+1}} d\exp(-\xi) \\
= & \int_0^1 \frac{\alpha}{u\left(1-\ln(u)+r.\left(-\ln(u)\right)^{\mu/\lambda}\right)^{\alpha+1}} du
\end{aligned}
$$

where

$$u = \exp(-\xi) \Leftrightarrow \xi = -\ln(u)$$

The latter integral can be approximated by

$$\frac{1}{m}\sum_{j=1}^{m} \frac{\alpha}{\frac{j}{m}\left(1-\ln(j/m)+r.\left(-\ln(j/m)\right)^{\mu/\lambda}\right)^{\alpha+1}}$$

where m is "large".

B.4 Interval Probability Distributions

Let

$$T_s \in [a,b], T_c \in [c,d],$$

and suppose that the intervals overlap, say:

$$a < c < b$$

Then

$$\begin{aligned}
& P\left[T_s \in [a,b], T_c \in [c,d]\right] \\
= \ & P\left[T_s \in [a,c] \cup [c,b], T_c \in [c,b] \cup [b,d]\right] \\
= \ & \int_a^c \left[\int_c^b f(t_c,t_s)dt_s + \int_b^d f(t_c,t_s)dt_s\right] dt_c \\
& + \int_c^b \left[\int_c^b f(t_c,t_s)dt_s + \int_b^d f(t_c,t_s)dt_s\right] dt_c \\
= \ & P\left[T_s \in [a,c], T_c \in [c,b]\right] \\
& + P\left[T_s \in [a,c], T_c \in [b,d]\right] \\
& + P\left[T_s \in [c,b], T_c \in [c,b]\right] \\
& + P\left[T_s \in [c,b], T_c \in [b,d]\right]
\end{aligned}$$

Now only

$$P\left[T_s \in [c,b], T_c \in [c,b]\right]$$

has overlapping intervals, but that is no longer a problem, because

$$\begin{aligned}
& P\left[T_s \in [c,b], T_c \in [c,b]\right] \\
= \ & S(c,c) - S(b,c) - S(c,b) + S(b,b)
\end{aligned}$$

is well-defined for both the control group and the the treatment group. Thus, any cartesian product $[a, b] \times [c, d]$ can be disentangled in unions of cartesian products of nonoverlapping intervals and common same intervals.

References

Abbring, J., and G. V. den Berg (2000a): "The Non-Parametric Identification of Treatment Effects in Duration Analysis," Unpublished Paper, Free University of Amsterdan.

Abbring, J., G. van den Berg, and J. van Ours (1997): "The Effect of Unemployment Insurance Sanctions on the Transition Rate from Unemployment to Employment," Working Paper, Tibergen Institute.

Abbring, J. H., and G. V. den Berg (2000b): "The Non-Parametric Identification of the Mixed Proportional Hazards Competing Risks Model," Manuscript, Free University of Amsterdam.

Angrist, J. (1998): "Estimating the Labor Market Impact of Voluntary MIlitary Service Using Social Security Data on MIlitary Applicants," *Econometrica*, 66, 249–288.

Arnold, B. C., and P. L. Brocket (1983): "Indentifiability for Dependent Multiple Decrement/Competing Risks Models," *Scandinavian Actuarial Journal*, 31, 117–127.

Barton, R. R., and B. W. Turnbull (1981): *A Failure Rate Regression Model for the Study of Recidivism*Models in Quantitative Criminology. Academic Press, New York, j. a. fox edn.

Beck, A. J., and B. E. Shipley (1989): "Recidivism of Prisoners Released in 1983," Special report, Bureau of Justice Statistics.

BECKER, G. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169–217.

BLUMSTEIN, A., AND J. COHEN (1979): "Estimation of Individual Crime Rates from Arest Records," *Journal of Criminal Law and Criminology*, 70, 561–585.

CARD, D., AND D. SULLIVAN (1988): "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56(3), 497–530.

CARR-HILL, G. A., AND R. A. CARR-HILL (1972): "Reconviction as a Process," *British Journal of Criminology*, 12, 35–43.

CHUNG, C.-F., P. SCHMIDT, AND A. WITTE (1991): "Survival Analysis: A Survey," *Journal of Quantitative Criminology*, 7(1), 59–98.

COPAS, J. B., AND F. HEYDARI (1997): "Estimating the Risk of Reoffending by Use of Exponential Misture Models," *Journal of the Royal Statistical Association*, A 160, 237–252.

COX, D. R. (1959): "The Analysis of Exponentially Distributed Lifetimes with Two Types of Failure," *Journal of the Royal Statistical Society*, B 21, 411–421.

——— (1961): "Test of Separate Families of Hypotheses," in *Proceedings of the Fourth Berkley Simposium on Mathematical Statistic and Probability*, pp. 105–123, Berkley. University of California Press.

——— (1962): "Futher Results on Tests of Separable Families of Hypotheses," *Journal of the Royal Statistical Society*, Seires B 24, 406–424.

——— (1972): "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, 34 B, 187–220.

——— (1975): "Partial Likelihood," *Biometrica*, 62(2), 269–276.

Crowder, M. J. (2001): *Classical Competing Risks*. Chapman and Hall - CRC.

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

Eberwein, C., J. Ham, and R. 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.

Ebrahimi, N. (1996): "The Effects of Misclassification of the Actual Cause of Death in Competing Risks Analysis," *Statistical Medicine*, 15, 1557–1566.

Ehrlich, I. (1973): "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," *Journal of Political Economy*, 81, 521–565.

Escarela, G., B. Francis, and K. Soothill (2000): "Competing Risks, Persistence, and Desistance in Analyzing Recidivism," *Journal of Quantitative Criminology*, 16(4), 385–414.

Farragi, D., and E. L. Korn (1996): "Competing Risks with Frailty Models when Treatment Affects Only One Failure Type," *Biometrika*, 83, 467–471.

Farrington, D. P. (1997): *Human Development and Criminal Careers*chap. 9, pp. 361–408, Handbook of Criminology. Claredon Press, Oxford, 2nd edn.

Farrington, D. P., and R. Tarling (1985): *Prediction in Criminology.* University of New York State Press.

Freuden, J. E. (1961): "A Bivariate Extention of the Exponential Distribution," *Journal of the American Statistical Association*, 56, 971–977.

Gendreau, P., T. Little, and C. Goggin (1996): "A Meta-Analysis of the Predictors of Adult Ofender Recidivism: What Works!," *Criminology*, 34(4), 575 – 607.

Gourieroux, C., and A. Monfort (1994): *Testing Non-nested Hypotheses*vol. IV, chap. 44, pp. 2585–2637. Elsevier Science, r. f. engle and d. l. mcfadden edn.

Gross, A. J., V. A. Clark, and V. Liu (1971): "Estimation of Survival Paramenters when One of Two Organs Must Function for Survival," *Biometrics*, 27, 369–377.

Gumbel, E. J. (1960): "Bivariate Exponential Distribution," *Journal of The American Statistical Association*, 55, 698–707.

Ham, J. C., and LaLonde (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Trainning," *Econometrica*, 64(1), 175–205.

Hausman, J. A., and D. A. Wise (1981): *Stratification on Endogenous Variables and Estimation: The Gary Income Maintenance Experiment*chap. 10, pp. 365–391, Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, c. f. manskin and daniel mcfadden edn.

Heckman, J. (1992): *Evaluating Welfare and Trainning Programs*chap. Randomization and Social Program. Harvard University Press, charles manskin and i. garfinkel edn.

——— (1996): "Randomization as an Instrumental Variable," *Review of Economics and Statistics*, 78(2), 336–341.

——— (1999): "Introduction to Selection Bias and Program Evaluation," University of Chicago.

Heckman, J., H. Ichimura, and P. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence fromEvaluating a Job Trainning Programme," *Review of Economic Studies*, 64, 605–654.

Heckman, J., R. J. Lalonde, and J. A. Smith (1999): *The Economics and Econometrics of Active Labor Market Programs*vol. 3A of *Handbook of Labor Economics*, chap. 31, pp. 1865–2097. Elsevier Science, Amsterdan.

Heckman, J., L. Lochner, and C. Taber (1998): "General-Equilibrium Treatment Effects," *American Economic Review*, 88, 381–386.

Heckman, J., and R. Roob (1985): *Longitudinal Analysis of Labor Market Data*chap. Alternative Methods for Evaluating the Impact of Interventions, pp. 156–245. Cambridge University Press, New York.

Heckman, J. J. (2000): "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *The Quarterly Journal of Economics*, pp. 45–97.

Heckman, J. J., and B. E. Honore (1989): "The Identifiability of the Competing Risks Model," *Biometrika*, 76(2), 325–330.

Heckman, J. J., and J. R. Walker (1987): *Using Goodness of Fit and Other Criteria to Choose Among Competing Duration Models: A Case Study of Hutterite*

*Data*chap. 9, pp. 247–287, Sociological Methodology. American Sociological Association, Wasington, D. C., c. cloog edn.

Hirano, K., G. Imbens, and G. Ridder (2000): “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” Manuscript, UCLA.

Hougaard, P. (2000): *Analysis of Multivariate Survival Data*. Springer.

ICPSR (1989): “Recidivism Among Released Prisoners, 1983,” Discussion Paper 8875, Inter University Consortium for Political and Social Research.

Imai, S., and K. Krishna (2001): “Employment, Dynamic Deterrence and Crime,” Discussion Paper 8281, NBER.

Imbens, G., and J. Angrist (1994): “Identification and Estimation of Local Treatment Effects,” *Econometrica*, 62, 467–476.

Katz, L. F., and D. B. Meyer (1990): “Unemployment Insurance, Recall Expectations and Unemployment Outcomes,” *Quartely Journal of Economics*, 105, 973–1002.

Kiefer, N. M. (1988): “Economic Duration Data and Hazard Functions,” *Journal of Economic Literature*, XXVI, 646–679.

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

Lattimore, P. K., C. A. Visher, and R. L. Linster (1994): “Specialization in Juvenile Careers: Markov Results for a California Cohort,” *Journal of Quantitative Criminology*, 10, 291–316.

LEHMANN, E. (1999): *Elements of Large-Sample Theory*. Springer, New York, first edn.

LEVITT, S. (1997): "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, 87(3), 270–290.

LILLARD, L. A. (1993): "Simultaneous Equations for Hazards," *Journal of Econometrics*, 56, 189–217.

LILLARD, L. A., AND C. W. PANIS (1996): "Marital Status and Mortality: The Role of Health," *Demography*, 33, 313–327.

MALTZ, M. D. (1984): *Recidivism*, Quantitative Studies in Social Sciences. Academic Press, Orlando, FL.

MALTZ, M. D., AND R. MCCLEARY (1977): "The Mathematics of Behavioral Change," *Evaluation Quarteely*, 3, 124–131.

MEYER, B. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58(4), 757–782.

MILKMAN, R. (2001): "Employment Services for Ex-Offenders, 1981-1984: Boston, Chicago, and San Diego," Discussion Paper 8619, ICPSR.

MOCAN, N., AND D. REES (1999): "Economic Conditions, Deterrence and Juvenile Crime: Evidence from Microdata," Discussion Paper 7405, NBER.

MOESCHBERGER, M. (1974): "Life Tests under Dependent Competing Causes of Failures," *Technometrics*, 16, 39–47.

MOFFITT, R. A. (1996): "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association*, 91, 462–465.

MYERS, S. (1983): "Estimating the Economic Model of Crime: Employment versus Punishement Effects," *Quarterly Journal of Economics*, pp. 157–166.

MYLKMAN, R., A. TIMROTS, A. PEYSER, M. TOBORG, B. G. A. YEZER, L. CARPENTER, AND N. LANDSON (1985): "Employment Services for Ex-Offenders Field Test," Discussion paper, The Lazar Institute.

PARTANEN, J. (1969): "On Waiting Time Distributions," *Acta Sociologica*, 12, 132–143.

PESARAN, M. H., AND M. WEEKS (1999): "Non-nested Hypothesis Testing: An Overview," Manuscript, University of Cambridge.

PISSARIDES, C. A. (2000): *Equilibrium Unemployment Theory*. The MIT Press, Cambridge, MA, second edn.

PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY, AND W. T. VETTERLING (1989): *Numerical Recipes:The Art of Scientific Computing (FORTRAN Version)*. Cambridge University Press, Cambridge.

PUDNEY, S. (1989): *Modelling Individual Choices: The Econometrics of Corners, Kinks and Holes*. Basil Blackwell, Oxford, first edn.

QUANDT, R. (1972): "Methods for Estimating Switching Regressions," *Journal of the American Statistical Association*, 67(338), 306–310.

RHODES, W. (1986): "A Survival Model with Dependent Competing Events and Right-Hand Censoring: Probation and Parole as Illustration," *Journal of Quantitative Criminology*, 2, 113–137.

ROSENBAUM, P. R., AND D. B. RUBIN (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 Treatments in Radomized and Nonrandomized Studies," *Journal of Education Psychology*, 66, 688–701.

SCHMIDT, P., AND A. D. WITTE (1988): *Predicting Recidivism Using Survival Models*, Research in Criminology. Springer-Verlag.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

STOLLMACK, S., AND C. M. HARRIS (1974): "Failure Rate Analysis Applied to Recidivism Data," *Operation Research*, 23, 1192–1205.

TIMROTS, A. (1985): "An Evaluation of Employment Services Programs for Ex-Offenders," Master's thesis, University of Maryland, College Park.

TSIATIS, A. (1975): "A Nonindentifiability Aspect of the Problem of Competing Risks," *Procedures of the National Academy of Science*, 72, 20–22.

VAN DEN BERG, G. (1997): "Association Measures for Durations in Bivariate Hazard Rate Models," *Journal of Econometrics*, 79, 221–245.

——— (1999): "Empirical Inference with Equilibrium Search Models of the Labor Market," *The Economic Journal*, pp. F283 – F306.

——— (2000): *Duration Models: Specification, Identification, and Multiple Durations*vol. V of *Handbook of Econometrics*. North Holland, Amsterdam.

van den Berg, G., and M. Lindeboom (1998): "Attrition in Panel Survey Data and the Estimation of Multi-State Labor Market Models," *The Journal of Human Resources*, XXXIII(2), 458–478.

van den Berg, G., M. Lindeboom, and G. Ridder (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., B. van den Klaauw, and J. C. van Ours (1998): "Punitive Sanctions and the Transition Rate from Welfare to Work," Working Paper, Tinbergen Institute, Amsterdan.

van den Berg, G., B. van der Klaauw, and J. van Ours (1998): "Punitive Sanctions and the Transition Rate from Welfare to Work," Working Paper, Tinbergen Institute.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979): "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality," *Demography*, 16, 439–454.

Visher, C. A., and R. L. Linster (1990): "A Survival Model of Pretrial Failure," *Journal of Quantitative Criminology*, 6, 153–184.

Witte, A. (1980): "Estimating the Economic Model of Crime with Individual Data," *Quarterly Journal of Economics*, pp. 57–84.

Wooldridge, J. (2001): "Asymptotic Properties of Weighted M-Estimators for Standard Startified Samples," *Econometric Theory*, 17, 451–470.

——— (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, first edn.

Zamble, E., and V. L. Quinsey (2001): *The Criminal Recidivism Process*. Cambridge University Press.

Vita

Jose Raimundo Carvalho was born in Fortaleza, Ceará, Brazil on March 2, 1968. In 1992 he received the B.Sc. in Civil Engineering, from Universidade Federal do Ceará, Brazil. In 1994, he received the M.Sc. in Economics from Universidade Federal do Ceará. Since 1995 he has been working as an assistant professor at Universidade Federal do Ceará, first in the Departamento de Teoria Econômica and then in the Departamento de Administração. In 1992 he married Maria Auxiliador Ferreira and, after a while, Cecília Carvalho was born. After four years, Sofia Carvalho was born. In 1998, by means of a fellowship from CAPES, Brazil, he enrolled the Ph.D. program in economics at the Pennsylvania State University. From 1998 to 2002 he was employed in the Economics Department of Pennsylvania State University as a teaching assistant and, lately, as a research assistant. Also, as part of his duties, he taught ECON 315, Labor Economics, five consecutive semesters for undergraduates at Pennsylvania State University. In October of 2002, he obtained his Ph.D. in Economics and returned to assume his duties at Universidade Federal do Ceará.