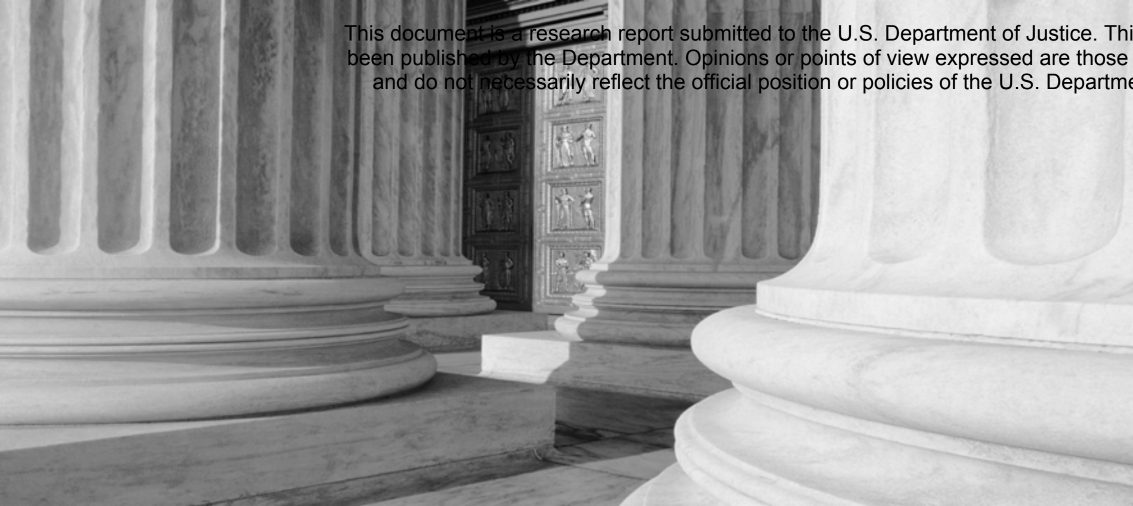
Document Title:	**Studying the Effects of Incarceration on Offending Trajectories: An Information-Theoretic Approach**
Author(s):	**Avinash Singh Bhati**
Document No.:	**216639**
Date Received:	**December 2006**
Award Number:	**2005-IJ-CX-0008**



TECHNICAL REPORT

July 2006

Studying the Effects of Incarceration on Offending Trajectories: An Information-Theoretic Approach

Technical Report Submitted to the Data Resources Program (DRP), National Institute of Justice (NIJ)

Avinash Singh Bhati

This report was prepared under Grant 2005-IJ-CX-0008 from the National Institute of Justice. Opinions expressed in this document are those of the authors, and do not necessarily represent the official position or policies of the U.S. Department of Justice, the Urban Institute, its trustees, or its funders.

research for safer communities

URBAN INSTITUTE
Justice Policy Center

Studying the Effects of Incarceration on Offending Trajectories: An Information-Theoretic Approach

Technical Report Submitted to the
Office of Justice Program
National Institute of Justice
U.S. Department of Justice

Avinash Singh Bhati
abhati@ui.urban.org / (202) 261-5329

Justice Policy Center, The Urban Institute
2100 M Street, N.W., Washington, D.C. 20037
July 31, 2006

Abstract

Research Goals and Objectives: The main goal of this project was to develop an analytical approach that will enable researchers, analysts, and practitioners to utilize detailed dated criminal history information, when such information is available, in order to investigate whether, and to what extent, incarceration is able to deter offenders from future offending. A secondary goal of the project was to demonstrate the utility of the developed framework by applying it to a real-world dataset.

Research Design and Methodology: The methodology developed in this project builds on two traditions. It uses concepts common to information-theory and event-history analysis. When combined, the resulting framework allows analysts (i) to estimate individual-specific offending micro-trajectories; (ii) to project counterfactual trajectories (i.e., trace out the offending trajectory for each individual had (s)he not been incarcerated); and (iii) to assess the actual post-release offending patterns against the backdrop of these counterfactuals. The information-theoretic underpinnings of the framework also help quantify the extent of deviation between the counterfactual and actual micro-trajectories for each individual. This composite statistic allows one to classify individuals' incarceration as having had a deterrent, an incapacitative or a criminogenic effect on them.

Research Results and Conclusions: Dated arrest histories of a sample of prisoners released from state prisons in 1994, collected by the Bureau of Justice Statistics and publicly archived at ICPSR (Study # 3355), were used to model these trajectories and study their deflection. Estimated models largely confirmed expectations. Upon release, being later in the offending sequence exerted an upward pressure on the risk path (trajectory) relative to what was anticipated and, all else being equal, being closer to prior offending activity exerted a downward pressure on the trajectory relative to the counterfactual. Moreover, a comparison of the counterfactual and actual offending patterns suggests that most releasees were either deterred from future offending (40%) or merely incapacitated (56%) by their incarceration. About 4% had a criminogenic effect.

Contents

List of Figures

List of Tables

Acknowledgements

This research was supported by grant # 2005-IJ-CX-0008 from the Data Resources Program of the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view expressed here are those of the author and do not represent the official positions or policies of the U.S. Department of Justice nor of the Urban Institute, its trustees or funders.

I thank Dr. Christy Visher of the Urban Institute for reviewing an earlier draft of this report and for providing very thoughtful comments and criticisms. Kevin Roland helped tremendously with the re-structuring of the data and with some of the analysis. Dionne Davis provided extremely capable editorial support. I thank Karen Beckman (formerly of the Urban Institute) with whom I had several discussions about this research effort during the early phases of this project. All errors and interpretations are, however, mine alone.

Executive Summary

BACKGROUND AND OBJECTIVES

Imprisonment, for any length of time, is a life-interrupting event. The process of reentry into society after a period of incarceration is ridden with questions of individual sustainability, vulnerability, and fear of failure. Therefore, identifying and understanding the effects that incarceration can have on different types of offenders under different contexts is crucial to developing strategies that minimize any criminogenic harm, and maximize any deterrent benefits, that result from it. This report describes an analytical framework designed to aid practitioners, analysts, and researchers in investigating these issues.

It builds on one of the well established and widely accepted empirical regularities in criminology: the link between an individual's past and future crime. Criminologists are not in complete agreement with regard to explanations of this link. However, none deny that such continuity in offending is a very real phenomenon. To the extent that such links exist, studying prior involvement in crime should provide useful insights into future offending patterns. This notion is validated in almost all studies of criminal recidivism—that prior criminal history is one of the best and most consistent predictors of recidivism.

It is also a well established fact in criminology that the rate of offending increases as youthful offenders age but that, at some point, the rate begins to decline. Hence, this non-monotonic shape (first increasing then decreasing)—termed the "age-crime curve"—is a very predictable aspect of offending over the life course. Given this second fact, it is not at all surprising that individuals' past involvement in crime predicts recidivism well. The total amount of crime accumulated by any individual at the time of release captures one aspect of the "age-crime curve." However, the second aspect of this relationship—the process by which individuals were accumulating their criminal histories—is seldom utilized in recidivism research in general, or for understanding the effects of incarceration in particular. Since it can be anticipated that individuals' involvement in criminal activities over the life course can be characterized (probabilistically) by a *trajectory*, then it should be helpful to study how incarceration deflects an individual's trajectory.

With this goal in mind, the objective of this research effort is to develop, and demonstrate the utility of an analytical framework that can aid practitioners, analysts, and researchers to:

- Model the pre-release criminal history accumulation process in order to characterize, as trajectories, the process by which these individuals had been accumulating their respective criminal histories;

- Use this knowledge as a way to project into the future what could reasonably have been expected of these individuals given their past—i.e., project a counterfactual trajectory; and

- Use this counterfactual trajectory as a backdrop against which to assess the actual post-release offending patterns.

The framework has the potential to help researchers answer a very basic question: *How does incarceration affect individuals?* This report describes one way of addressing this important question in terms of whether, and to what extent, incarceration is able to *deflect* the trajectory a particular offender is on. In order for any analytical framework to provide meaningful insights into this question it must confront three related problems. First, it needs to be able to model individuals' trajectories using knowledge of their past offending patterns. Second, it needs to be capable of projecting trajectories into the future. Finally, it needs to have a mechanism by which to compare actual and counterfactual trajectories for each and every individual so that their incarceration can be appropriately classified as having had a deterrent, a criminogenic, or an incapacitative effect on them.

The information-theoretic approach described in this report is one approach that offers each of these capabilities. It only requires that detailed dated arrest histories, both before incarceration and after prison release, be available to the analyst. Moreover, it provides the usual statistical inferential apparatus whereby analysts can gauge the sensitivity of their results to sampling variation—i.e., how different their estimates would be had a slightly different sample been used. The report provides detailed derivations of the analytical framework and points readers to appropriate sources in the related econometrics/statistics literatures.

DATA USED

The developed framework is tested using a real world data set. In early 2002, the Bureau of Justice Statistics issued a report titled *Recidivism of prisoners released in 1994* that reported on criminal re-involvement of a sample of roughly 38,000 prisoners who were released in 1994 from prisons in 15 states (Langan and Levin, 2002). The data used to support their findings were subsequently archived at the National Archive of Criminal Justice Data at the Inter-University Consortium of Political and Social Research (study # 3355). These data contain detailed information on up to 99 arrest events for each of the individuals in the sample. This includes their pre-incarceration arrest events as well as arrest events within a period of 3 years after release. In addition, the data provide standard demographic information on each of the individuals as well as some limited information on their 1994 release mechanism.

To show how the developed framework may fruitfully be applied by researchers, analysts, and practitioners having access to such detailed data the BJS recidivism data were used as a test bed. The report describes in detail how the data were restructured, what predictable patterns were found in the data, and provides detailed estimates of the models. Once modeled, the counterfactual trajectories of each individual in the sample were compared with the actual post-release offending patterns in order to classify the effect that incarceration had in deflecting these trajectories. Finally, the limited set of explanatory information available in these data were used to model and study what factors, if any, helped explain the kinds of experiences people were expected to have. Unfortunately, this source provides insufficient data to make sound policy recommendations about what factors (or policy options) can be expected to maximize the deterrent benefits (or minimize criminogenic harm) of incarceration. The results presented in this report, for this part of the analysis, are intended primarily to showcase the capabilities of the developed framework.

FINDINGS

Despite the emphasis of this research effort being on the *development* of the framework, some interesting findings are summarized below.

- There was a fair amount of consistency among all the pre-prison based models of the criminal history accumulation processes across the 15 states analyzed. For example, being further along in the criminal career (i.e., being at risk of a higher arrest number) and starting the career later (i.e., having a higher age at first arrest) are consistently associated with lowered hazard trajectories. Similarly, all else being equal, being closer to past arrest clusters is consistently associated with an increased hazard trajectory. There was less consistency among states when modeling the deviation between the counterfactual and actual rearrest trajectories after release. Being later in the criminal career was found to exert an upward pressure on the offending trajectory relative to the counterfactual. Similarly, being closer to past cluster was found to exert a downward pressure on the trajectory relative to the counterfactual.

- The criminal history accumulation process contained valuable information about the long-term trends in individuals' offending patterns over the life course. The counterfactual trajectories, based on estimated models of the pre-prison based criminal history accumulation process and projected for the post-release period, perform remarkably well in predicting rearrests within three years of release. On the other hand, these same counterfactuals do not perform as well when used for making short-term projections. The false-positive rates are at very high levels throughout the follow-up period. When updated with models of the post-release behavior, the models perform much better.

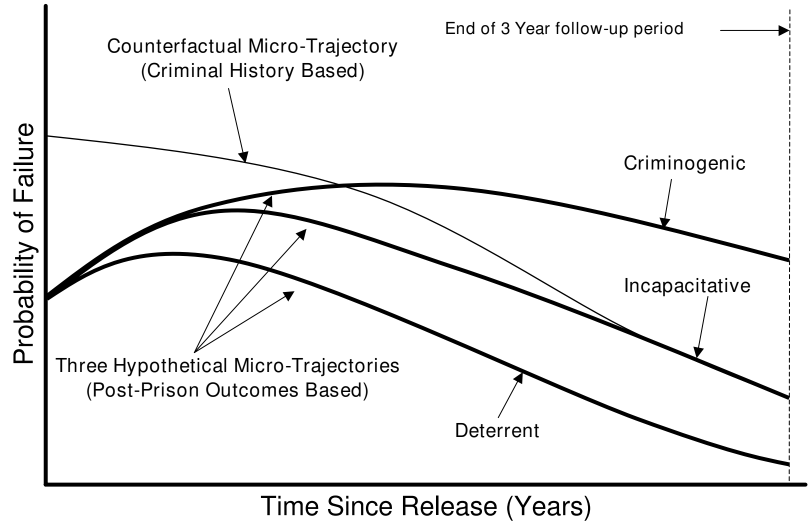


Figure A: A counterfactual trajactory compared with three hypothetical post-release trajectories showing criminogenic, incapacitative, and deterrent effects of incarceration.

- Information-theoretic measures were developed to quantify and classify the divergence between the counterfactual and the actual post-release micro-trajectories. Figure A displays three hypothetical post-release trajectories compared to a counterfactual and how each would be classified. Based on those computations, and in this analysis, large portions of the released cohort were classified as having had an incapacitative (56%) or a deterrent (40%) experience. A small proportion of the sample (4%) experienced criminogenic effects as a result of this incarceration.
- Using these classifications as the criterion outcome, being older at release and being closer to past clusters were consistently found to *increase* the likelihood of a releasee being deterred. Having more prior accumulated arrests and having a later age at first arrest were both found to significantly *decrease* the likelihood of a deterrent effect. Being released to supervision was found not to deter releasees substantially.
- Using the average log divergence between the counterfactual and the actual trajectories as the criterion some conflicting findings emerged. However, the effects of age at first arrest and age at release were qualitatively similar to what were found in the categorical analysis. Additionally, females experienced larger deterrent effects compared to similar males.

IMPLICATIONS

This research effort has important substantive, methodological, and practical implications.

- *Substantive implications*. Substantively, the analytical framework developed here has the potential to shed light on a very important question: *How does incarceration affect individuals?* The framework allows researchers to determine, or at the very least investigate, the types of individuals likely to be deterred by incarceration. In a similar way, it allows them to better understand how incarceration can have differing impacts on the same people at various stages in their life and/or criminal careers.

- *Methodological implications*. When the detailed dated arrest histories of a sample of releasees is available to researchers, utilizing only one source of variation in the data—the total amount of criminal history accumulated prior to prison admission—when modeling the risk of future recidivism forces analysts to waste valuable information and thereby forgo learning opportunities. A second source of variation available in these pre-prison arrest histories—the process by which individuals were accumulating these histories—contains immense amount of information about future offending patterns. The information-theoretic event-history models, developed in this research effort, show how this knowledge can be introduced into the modeling strategy in a very effective way. The process by which individuals accumulate their pre-prison arrest histories, typically, have very predictable patterns that can be modeled. These models allow projection of person-specific micro-trajectories that trace out the evolution of rearrest risk *had the individual not been incarcerated*. As such, they are perfect counterfactuals against which to assess post-release offending patterns.

- *Practical Implications*. Although much of the software needed for the analysis conducted here needed to be programmed from scratch, the availability of standard software allowing researchers to utilize information and entropy based methods is increasing rapidly. For example, SAS has introduced an experimental procedure under its ETS module called PROC ENTROPY that is designed for the estimation of linear and non-linear models using the Generalized Maximum Entropy (GME) approach introduced by Golan, Judge, and Miller (1996). Additionally, LIMDEP—another popular econometrics software—has recently added the GME methods for estimating binary and multinomial logit models.

  Software needed to estimate generalized hazard models using the framework described in this report here is far from being developed. In the interim, researchers and practitioners will need to rely on routines and macros developed and made available to the public. An Appendix to this report provides a sample SAS program that was developed to estimate the models presented in this report.

FUTURE RESEARCH

As a result of this research effort, and based on the findings reported in this report, some recommendations for future research can be enumerated.

- The emphasis in this research effort was on development of the analytical framework and demonstration with an application. Comparison of the developed framework to existing and related approaches remains to be done as does the work of assessing the framework's performance using artificially generated data. Such simulation exercises are crucial to establish the credibility of the modeling approach as well as its performance relative to others.

- The framework can also be fruitfully extended to study the trajectories of multiple types of repeatable events such as offending and drug use over the life-course, or offending and employment, etc. Such analyses have the potential of shedding light on how incarceration can interrupt the *co-evolution* of these interrelated behaviors.

- The framework can also be extended to study how other interventions, not just incarceration, may deflect the trajectories of offending. For example, the effects of participation in various treatment programs may be quantified in terms of the program's ability to deflect individuals' offending trajectories.

Chapter 1

Background and Motivation

Imprisonment, for any length of time, is a life-interrupting event. The process of reentry into society after a period of incarceration is ridden with questions of individual sustainability, vulnerability, and fear of failure. Therefore, identifying and understanding the effects that incarceration can have on different types of offenders under different contexts is crucial to developing strategies that minimize any criminogenic harm, and maximize any deterrent benefits, that result from it. This report describes an analytical framework designed to aid practitioners, analysts, and researchers in investigating these issues.

One of the well established and widely accepted empirical regularities in criminology is the link between an individual's past and future crime.[1] Criminologists, however, are not in agreement as to the explanation for this persistence in and, more interestingly, divergence from criminal behavior. Due in large part to the publication of the 1986 National Academy of Sciences report on criminal careers and career criminals (Blumstein, Cohen, Roth, and Visher, 1986), the last three decades have seen a surge in research activity that has sought to theorize and explain continuity and change in crime as well as to inform policy of the appropriate role incarceration can and should play in crime control.

The theoretical debate, and the related empirical debate, centers on the causal interpretation attributed to the link between past and future crime. Some criminologists argue that this link is simply a manifestation of a constant (unchanging) criminal propensity, where others argue that the link between past and future crime is causal. The policy relevance of this debate is obvious: To the extent that an individual's relative criminal propensity is "fixed", incarceration can and should play only an incapacitative role. If, on the other hand, an individual's relative criminal propensity is not "fixed", then incarceration could serve as a deterrent and possible turning point to desistance from crime. See, among others, Hirschi and Gottfredson 1983; Farrington 1986; Sampson and Laub 1990,1993; and Piquero, Farrington, and Blumstein 2003 for reviews of the theoretical and methodological

[1]Over two-thirds of individuals released from prison nationwide, for example, were rearrested for a new crime within three years of release (Langan and Levin 2002). When attempting to explain such high recidivism rates, researchers typically find that releasee's criminal histories are the most reliable predictors.

issues surrounding the criminal career paradigm.

As a point of departure, this research effort acknowledges the possibility that incarceration could have an incapacitative, a deterrent, or a criminogenic effect on *every* releasee. In this report, I explain and demonstrate the utility of applying an event history-based, information-theoretic method for modeling the detailed criminal history accumulation process of every releasee and, furthermore, for using this process as a backdrop against which to analyze and understand the releasee's post-prison offending behavior.

Although the developed analytic framework has multiple uses, here I have utilized it for a very specific purpose: to compare every releasee's post-prison and pre-prison offending micro-trajectories in order to assess whether *this* incarceration episode had an incapacitative, a deterrent or a criminogenic effect on each of the releasees in the sample. Theoretically, these classifications can then be linked to individual, contextual, and policy relevant variables in an attempt to understand what factors are related to these three types of experiences. The research effort was aimed mainly at developing the analytical framework and not at providing any specific policy recommendations. Fortunately, sufficiently detailed data were available for the first part of the analysis—i.e., modeling the detailed criminal history accumulation process and comparing the pre-release and post-release micro-trajectories. The detailed explanatory data needed for the latter half of the analysis, however, were not available. Therefore, only a limited set of results are presented and discussed here in order to demonstrate how this approach may be helpful to practitioners.

This research effort builds on prior research on recidivism, generally, and post-prison recidivism research, specifically, although the emphasis is different.[2] Its goal was not to develop models of recidivism as prediction tools (per se). Rather, its goal was to develop tools for estimating and comparing a releasee's actual post-prison offending trajectory with (her)his criminal history-based *counterfactual* offending trajectory for the sole purpose of answering the question: "How, if at all, has this incarceration experience *deflected* the trajectory (path, career) the offender was on?" Since the offender, in question, was incarcerated and had (her)his career interrupted, the pre-prison offending micro-trajectory is termed a *counterfactual* because we never actually observe what this individual would have done had (s)he not been incarcerated. The strategy developed here is a flexible way of using *all* available knowledge about prior offending patterns to make inferences about post-prison offending trajectories.

This idea is not new. Bushway, Brame, and Paternoster (2004:97), for example, note that "... [P]re-existing rates of offending at the time of incarceration would be a perfect

[2]There exists a significant literature on modeling criminal (or other) recidivism using fully- and semi-parametric survival-type duration models from single or multiple (split) populations. Maltz (1984) and Schmidt and Witte (1988) are authoritative early texts on this topic. More recently, researchers have begun to link these approaches with the study of desistance from criminal careers (Brame, Bushway and Paternoster, 2003; Bushway, Brame, and Paternoster, 2004) using probabilistic definitions of desistance. The aim of this research effort is to develop tools for understanding the effects of incarceration on post-release offending behavior, classify these effects, and investigate their correlates.

control for individual heterogeneity." But, two individuals with exactly the same pre-incarceration offending rates may have been on differently sloped trajectories at the time of incarceration and, given varying lengths of time served in prison, could be released at very different times in their lives/careers. The analytical strategy developed here, in utilizing a projected counterfactual for *each and every* individual, is a flexible and robust means of explicitly taking these differences into account.

The methodological challenge lies with developing this counterfactual and in assessing whether, and to what extent, the (actual) post-prison offending trajectory deviates from the counterfactual, and subsequently classifying the incarceration experience accordingly.

Identifying and understanding the correlates of these distinct experiences should be of tremendous help to correctional authorities in reentry planning. Knowledge about the types of releasees likely to experience criminogenic or deterrent effects as a result of their incarceration, for example, could be used in the development of support systems designed to foster positive reentry experiences, and could be a crucial ingredient to individual successes, and ultimately to the promotion of public health and safety.

Crucial to the proposed analytic approach is the availability of the calendar dates of the criminal events that constitute an individual's past criminal record, as well as the dates of post-release criminal events within the follow-up period. This project relied on the recently released BJS study documenting the detailed criminal histories (as measured by arrests) of a sample of approximately 38,000 offenders released from state prisons across 15 states in 1994. These data were collected by BJS and are publicly available at NACJD (ICPSR). Unfortunately, these data do not provide the kind of detailed information that would be needed to make recommendations regarding specific policy options that may affect the likelihood of incarceration being a deterrent (rather than merely incapacitative or actually criminogenic). To the extent that state and local authorities have access to such detailed data, the analytical framework explained in this report can be applied in a straight forward manner.

Chapter 2

The Analytical Framework

As noted in the previous chapter, the basic challenge researchers or practitioners face in assessing the effects of incarceration on a particular releasee is two-fold. First, they must decide on an appropriate measure that quantifies the outcome they wish to assess. For example, they may wish to assess the effects of incarceration on the risk of post-release relapse into crime, chances of post-release employment, or the risk of relapse into drug use, etc. Having decided on an appropriate outcome, they must then develop plausible counterfactuals for the post release period. The effects of incarceration can then be assessed using this counterfactual. In this report, I restrict attention to the risk of recidivism (as measured by rearrest) as the outcome of interest. Extensions to other types of outcomes and/or multiple outcomes are possible and left for future work.

The challenge of developing plausible counterfactuals then boils down to developing estimates of each individual's risk of recidivism for the follow-up period *had they not been incarcerated*. If such an estimate can be developed then one can compare each releasee's actual post-release offending behavior to this counterfactual and use this as a way to classify the releasee's incarceration experience. Incarceration can be classified as having had a deterrent, incapacitative, or a criminogenic effect on a releasee depending on whether his(her) risk of recidivism is found to be lower, about the same, or higher than the counterfactual.

Unfortunately, however, the risk of recidivism is not a static but a dynamic measure. Quantification of the risk of recidivism (statement about "how much") must be accompanied, implicitly or explicitly, by statements about "when". For example, a statement like "person A's risk of recidivism is 20%" says little without the qualification that this pertains to a 2 year follow-up period after release. Therefore, what is needed is a re-definition of the outcome of interest as well as its counterfactual in terms of a dynamic *function* quantifying the evolution of the risk of recidivism over time rather than a static measure. Fortunately, techniques for the analysis of duration data offer a variety of ways of linking the risk of recidivism with time or age thereby allowing the estimation of this dynamic outcome. The remaining challenges then are to (a) develop a dynamic counterfactual micro-trajectory for each individual in the sample, and (b) develop ways to test for differences between the actual and counterfactual micro-trajectories. By comparing dynamic outcomes—the micro-trajectories—we would in fact be comparing whether incarceration has altered or deflected the trajectory (the career path) a particular releasee was on.

In this chapter, I explain an information-theoretic approach that can be used for (i) developing these counterfactual micro-trajectories utilizing detailed information about past arrest patterns and (ii) testing whether or not the post-release trajectory is, in some sense (to be described later), better, worse, or about the same as the counterfactual. Therefore, the effects of incarceration are classified based on whether or not incarceration has deflected "sufficiently" an individual from his(her) own counterfactual and, if so, whether this deflection is for the better or the worse in terms of the outcome of interest.

The chapter is organized as follows. I begin by developing information-theoretic models of offending trajectories using detailed dated arrest records of a group of offenders. These models can be applied to retrospective (historical) data as well as prospective sequences of events. The dated arrest histories allow detailed models of the risk of each successive arrest number (e.g., the first, second, third, and so forth) at all ages. Once estimated using retrospective criminal histories prior to prison admission, these models then allow projection of the rearrest risk trajectories for each individual given their age at release and the rearrest number they were then at risk of. These projections form the counterfactuals against which the actual rearrest patterns (post-release) can be assessed. Finally, I develop the tests of the divergence between the actual and counterfactual micro-trajectories.

The following conventions will be used throughout this report. Scalar quantities will be denoted by italicized letters (x_n) or greek symbols (β_k) with appropriate subscripts. Column vectors will be denoted by bold unitalicized letters ($\mathbf{x}_n$) or symbols ($\boldsymbol{\beta}_k$), again with appropriate subscripts as needed. Row vectors will be denoted with the transpose of the column vectors (e.g., $\mathbf{x}'_n$). Finally, matrices, where needed, will be denoted by bold unitalicized and capitalized letters ($\mathbf{X}$) and symbols ($\boldsymbol{\Phi}$). How scalars are gathered to construct vectors and how vectors are gathered to construct matrices will be made explicit when the relevant quantities are defined.

2.1. A SIMPLE NON-PARAMETRIC MODEL

Consider, as a point of departure, the following problem. We have available detailed dated sequences of events (arrests) for a group of individuals. To be concrete, I will restrict the explanation and discussion to arrest sequences although the models are just as applicable to other events. Also, the sequence can be prospective or retrospective histories of a particular cohort. Here, I will first develop the framework for the retrospective histories of arrest events prior to prison admission. The cohort of interest, therefore, is a sample of prisoners released at a particular time. For example, the cohort of interest for the application discussed in this report will be a sample of prisoners released from state prisons in 1994. It is assumed that detailed information pertaining to the pre-prison arrest histories are available for each of the individuals in the sample in addition to dated re-arrest event(s) within a finite window after the current release.

Detailed information pertaining to each arrest need to include, at a minimum, the date of the arrest and its order in the sequence (i.e., arrest number 1, 2, 3, etc.). Detailed information pertaining to the individuals need to include, at a minimum, the date of birth of the individual. This minimal amount of information is needed in order to construct sequences of ages at each successive arrest events. Harding and Maller (1997) refer to this sequencing as individuals' *arrest profiles*. Assume that such profiles exist for the period before incarceration and for a fixed period after release.

Given that a prison release cohort is likely to have variation in the age at release *and* variation in the amount of time served in prison, it can be expected that this cohort will have had varying amounts of time between their birth and the last prison admission (from which they are released in 1994). Therefore, we can expect to have available two sources of variation in the data. First, we can expect sufficient variation among the individuals with respect to the number of arrests accumulated prior to prison admission—i.e., the "amount" of criminal history accumulated. Second, we can expect variation in the way these arrest histories were accumulated—i.e., the criminal history accumulation "process." In most criminal recidivism research, the total *amount* of criminal history accumulated prior to release is a very strong determinant of future arrest risk. However, with few exceptions, researchers typically do not utilize the full variation in the criminal history accumulation *process* when assessing future rearrest risk.[1] In the analytical approach developed next, I make full use of this second source of variation.

First, some definitions. Let a_{rn} denote the age of the nth individual when (s)he was arrested for the rth time. The subscript $n = 1, \ldots, N$ is used to index individuals and $r = 1, \ldots, R_n$ is used to index arrest events. Each individual can have a different number of total arrests in the sequence (hence the limit R_n). Let us restrict, for the moment, the derivation only to the pre-release portion of the arrest profiles. This means we do not have to deal with censoring—the last arrest in each individual's sequence was what got them into prison for the R_nth time. After that, they were not at risk of any more arrests.

Next, let us artificially discretize the continuous "age at arrest" variable. That is, for M mutually exclusive and exhaustive artificially defined intervals (say monthly, quarterly, etc.), let us define the following dummy variables

$$y_{rmn} = \begin{cases} 1 & \text{if } a_{rn} \in (z_{m-1}, z_m) \\ 0 & \text{otherwise} \end{cases} \qquad \forall n \in N; r \in R_n; m \in M. \tag{2.1}$$

[1] However, there are some exceptions. Visher, Lattimore, and Linster (1991), for example, apply declining weights to prior arrest events thereby giving more relevance to arrest events in the recent past and lower relevance to arrest events from the more distant past. This allows them to develop a more refined criminal history score measure that they then use to model/predict future crime. However, this score is still a static measure that does not allow one to compute a counterfactual offending trajectory against which to assess post-release behavior. In fact, any score developed by a weighted or unweighted combination of prior arrest events can only provide a static measure and cannot be used to construct a dynamic counterfactual.

Table 2.1: An example of creating the y_{rmn} and d_{rmn} flags from arrest profiles.

			z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9
			0	5	10	15	20	25	30	35	40
n	r	a_{rn}	y_{r1n}	y_{r2n}	y_{r3n}	y_{r4n}	y_{r5n}	y_{r6n}	y_{r7n}	y_{r8n}	y_{r9n}
1	1	19	0	0	0	0	1	0	0	0	0
1	2	25	0	0	0	0	0	1	0	0	0
2	1	17	0	0	0	0	1	0	0	0	0
2	2	23	0	0	0	0	0	1	0	0	0
2	3	37	0	0	0	0	0	0	0	0	1
n	r	a_{rn}	d_{r1n}	d_{r2n}	d_{r3n}	d_{r4n}	d_{r5n}	d_{r6n}	d_{r7n}	d_{r8n}	d_{r9n}
1	1	19	1	1	1	1	1	0	0	0	0
1	2	25	0	0	0	0	1	1	0	0	0
2	1	17	1	1	1	1	1	0	0	0	0
2	2	23	0	0	0	0	1	1	0	0	0
2	3	37	0	0	0	0	0	1	1	1	1

In effect, we are creating a set of M binary dummy variables for each arrest event for each individual at each age.[2] Consider, next a positive quantity, denoted s_{rmn}, that we believe this set of dummy variables represent. We can think of the actual outcomes as a noisy (imperfect) manifestation of some underlying reality that we wish to recover. Given the assumption of imperfection, we can only link these unknown quantities (s_{rmn}) to their observed counterparts (y_{rmn}) as approximations. Therefore, let

$$y_{rmn} \approx s_{rmn} \qquad \forall r, m, n. \tag{2.2}$$

So far we have assumed that each event is a distinct outcome without regard to their order. To build in the order of the events we need to define a corresponding set of dummy variables that flag whether or not a particular event is possible at a particular age. Let

$$d_{rmn} = \begin{cases} 1 & \text{if } z_m \in (a_{(r-1)n}, a_{rn}) \\ 0 & \text{otherwise} \end{cases} \qquad \forall n \in N; r \in R_n; m \in M. \tag{2.3}$$

Here, unlike (2.1), we are creating a set of dummy variables flagging the possibility of each arrest event for each individual at each age. An example of what these dummy variables would look like for two arrest profiles is given in Table 2.1.

[2]Note, this is only for developing the model. As will be explained below, the artificial discretization of the continuous variable will be removed and the full variation in the continuous age will be used.

Individual 1, for example, was arrested for the first time at age 19 and for the second time at age 25 after which this individual entered prison and was released as part of the 1994 release cohort. Therefore, $y_{rmn} = y_{1,5,1} = 1$ and $y_{1,m,1} = 0 \ \forall m \neq 5$. Similarly, $y_{rmn} = y_{2,6,1} = 1$ and $y_{2,m,1} = 0 \ \forall m \neq 6$. Lets turn next to the d flags. For the first event, $d_{rmn} = d_{1,m,1} = 1 \ \forall m \leq 5$ and are set to 0 $\forall m > 6$. This is because the individual is not at risk of being arrested for the first time after (s)he has been arrested for the first time. The individual is now at risk of being arrested for the second time, i.e., $d_{rmn} = d_{2,m,1} = 1$ $\forall m \in (5, 6)$, until (s)he is arrested for the second time. After that the individual enters prison for the last time before being released in 1994.

Having defined the two interrelated sets of dummy variables, let us combine them. To do so, let us pre-multiply both sides of (2.2) by the d_{rmn} flags, sum across all individuals with the same r and m, and assume that this aggregation washes out all the imperfections. In other words, even though each y_{rmn} are only imperfect manifestations of the corresponding s_{rmn}, let their sums within r and m be perfectly preserved. This allows us to convert the inequalities into the following equalities:

$$\sum_n d_{rmn} y_{rmn} = \sum_n d_{rmn} s_{rmn} \qquad \forall r, m. \tag{2.4}$$

Finally, if we assume that $s_{rmn} = s_{rm} \ \forall r, m$, i.e., that this quantity is fixed within each r and m pairs, then we can solve explicitly for each of these unknown quantities to get

$$s_{rm} = \frac{\sum_n d_{rmn} y_{rmn}}{\sum_n d_{rmn}} \qquad \forall r, m. \tag{2.5}$$

Since an event occurs (i.e., $y_{rmn} = 1$) only when an individual is at risk of that event occurring (i.e., $d_{rmn} = 1$), we see that the numerator of this ratio is merely the number of individuals being arrested for the rth time within the mth age interval. The denominator, on the other hand, is merely the number of persons that were at risk of being arrested for the rth time during the mth age interval. This quantity is, of course, a familiar one. In statistics and econometrics it is referred to as the hazard (rate) and in demography it is referred to as a Parity Progression Ratio (PPR).[3] The derivation in (2.5) is in fact a nonparametric estimate of the hazard of the rth event occurring during age interval m (or the probability of progressing to the next event, conditional on being at risk of that progression).

Visually, this concept can best be explained by means of the Lexis diagram in Figure 2.1, where the criminal history accumulation process of five hypothetical offenders are shown.[4] Each diagonal line represents a releasee's life prior to the current incarceration. The filled black circles represent arrest events (with the arrest numbers indicated alongside them), and a hollow circle represents the arrest that resulted in the current incarceration.

[3] See Chapter 9 in Hinde (1998) for a general discussion of PPRs. See Feeney and Yu (1987) and Bhrolchain (1987) for applications of PPRs to changes in fertility patterns.

[4] See the Maltz and Mullany (2000) for other interesting ways in which this information could be visualized.

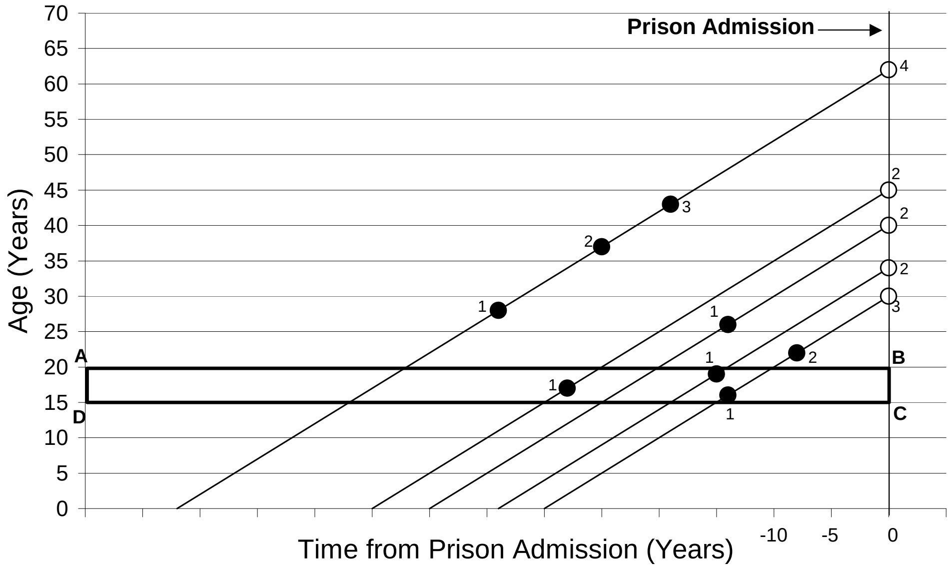


Figure 2.1: Lexis Diagram showing the Criminal History Accumulation Process of five hypothetical offenders entering prison

Now consider the rectangular region (ABCD). This region contains the criminal activities of the five hypothetical offenders during their 15 to 20-year life-phase. Since there were no criminal activities (for this group) prior to that age, we find that all five persons are at risk of initiating their criminal career at age 15 (i.e., $\sum_n d_{rmn} = 5$). Next, we note that, of these, only three actually committed their first crime during this phase of their life (i.e., $\sum_n d_{rmn} y_{rmn} = 3$). This allows us to compute the PPR for the first progression (initiation) by age 20 for this group of people as $s_{1m} = 0.6$. In a similar manner, we may compute the PPR for the first progression during other life-phases, and we may compute the PPR for subsequent progressions during this and other life-phases. In essence, for any given sample, we may use detailed dated criminal histories to construct PPRs that characterize the sample members' criminal history accumulation process.

The point of this analytical and graphical derivation was simply to demonstrate that, when combined with the set of dummy variables d_{rmn}, any manipulation of the left and right hand sides of (2.2) will yield constraints on the values the hazard can take. In fact, this same derivation can be extended without change of notation to include censored cases. We only need to define d_{rmn} to be 0 after the event has occurred *or* if the individual is no longer being observed, i.e., the case is censored. Note that for censored cases y_{rmn} will always be 0 because we never see the event occurring. However, between the $r-1$st event and the time of censoring, this individual will contribute towards the denominator of (2.5).

In the example provided in Table 2.1, I used 5-year intervals. In fact, one can use as small an interval as one desires. For example, when studying age profiles as measured in year, one can define intervals as small as a quarter or a month. However, computation of the nonparametric PPR or hazard becomes more unstable because we end up with many empty cells and many cells where the denominator is 0. This suggests moving towards a parametric formulation of the problem which allows a flexible functional form linking the hazards across persons, ages, and event numbers. I turn to that formulation next.

2.2. A SEMI-PARAMETRIC INFORMATION-THEORETIC APPROACH

Instead of assuming that $s_{rmn} = s_{rm}\ \forall n$, suppose we allow the hazard to vary across n, m, as well as r. What we need now is some way to impose structure on the model. Consider the minimal set of variables we have available in Table 2.1: z_m = the age grid points just defined, and r = the event number. A simple way to impose structure on the model is to weight both sides of the approximation (2.2) with z_m and r and take the weighted and unweighted sums across all n. This yields the following two equations:

$$\sum_{rn}\sum_{m} z_m d_{rmn} y_{rmn} = \sum_{rmn}\sum_{m} z_m d_{rmn} s_{rmn} \tag{2.6}$$

$$\sum_{rn} r \sum_{m} z_m d_{rmn} y_{rmn} = \sum_{rmn} r \sum_{m} z_m d_{rmn} s_{rmn} \tag{2.7}$$

Unfortunately, unless we make some assumptions about s_{rmn}, we cannot proceed to solve for their values like we did in the non-parametric case. However, under the Information-theoretic approach, to be developed below, we *can* recover information about the s_{rmn} without making any a-priori assumptions about the form of s_{rmn}.

Before proceeding to that, however, I generalize the problem to include an expanded set of explanatory variables that can vary across individuals, events, and time. As in all moment based methods, however, it is still assumed that the variables in this set are not *perfectly* correlated.

2.2.1. Setting up the basic problem

Suppose there exist a set of K event- and person-specific attributes (denoted x_{krn}) that we believe are part of the mechanism generating the outcomes—i.e., part of the hazard model. Minimally, this would include r and z_m as shown above. How do we introduce these attributes into the model? As explained in the previous section, introducing the order of event was accomplished simply by pre-multiplying both sides of (2.2) by the flags d_{rmn}. Introducing attributes can be done in much the same way, as explained above. That is, we can pre-multiply both sides of (2.2) by the corresponding d_{rmn} flags, the artificial discrete support points z_m, as well as the available attributes x_{krn}, and aggregate across m, r, and n in order to convert the inequalities into equalities. This yields the following K equality constraints:

$$\sum_{rn} x_{krn} \sum_{m} z_m d_{rmn} y_{rmn} = \sum_{rn} x_{krn} \sum_{m} z_m d_{rmn} s_{rmn} \qquad \forall k \in K. \tag{2.8}$$

where, we assume that the attributes includes a column of 1's. Note that each event can occur only once in an individual's lifetime (e.g., the 4th arrest can only occur once) and since we are dealing with the uncensored records (pre-prison criminal histories) we can assume that none of the at-risk periods end without an event. In other words, $\sum_m z_m d_{rmn} y_{rmn} \approx a_{rn}$ $\forall r, n$ on the left hand side (LHS) of (2.8).[5] But the approximation is merely a result of our artificial discretization of the continuous age variable and by making the discrete intervals arbitrarily small we, in fact, approach the continuous variable. Therefore, we can replace the term $\sum_m z_m d_{rmn} y_{rmn}$ on the LHS with the actual continuous age variable in order to utilize the full variation available to us. This allows us to re-write these constraints as:

$$\sum_{rn} x_{krn} a_{rn} = \sum_{rn} x_{krn} \sum_{m} z_m d_{rmn} s_{rmn} \qquad \forall k \in K \tag{2.9}$$

We now have, what is termed, an ill-posed inversion problem—more unknowns than equations linking them (Levine, 1980). In the non-parametric hazard model case, we solved this problem by assuming that $s_{rmn} = s_{rm}$ $\forall n$ and we only summed within each r and m pairs. In other words, we had exactly the same number of unknowns as we had constraints. That allowed us to explicitly solve for a particular solution by *inverting* the quantity multiplying each s_{rm} on the right hand side (RHS) and taking that to the LHS. Here, the problem is ill-posed. We have far more unknowns than we have constraints. How do we solve this ill-posed problem?

2.2.2. Information Theory: A brief digression

Edwin Jaynes (1957a; 1957b), in a series of influential papers in statistical physics proposed a solution to such a problem provided that the unknown quantities are in the form of proper probabilities. He proposed that when faced with a problem that has possibly an infinite number of solutions, we should choose the one solution that implies maximum uncertainty while ensuring that the constraints (evidence) are satisfied. That way, we will be making the most conservative (safe) use of the evidence. Jaynes (1982) provides axiomatic derivation of the rationale underlying this approach.

Of course, for it to be operationalized, we need some quantification of uncertainty. Within the context of a problem in communication theory, Shannon (1948) defined the uncertainty contained in a message with J mutually exclusive and exhaustive outcomes as $H = -\sum_j p_j \log p_j$ and termed it *Information Entropy*. Here p_j is the probability that we will observe event j from the set of J possible events. In what came to be known as

[5] Extending this to censored cases is trivial and will be discussed later.

the Maximum Entropy formalism (or the principle of insufficient reason), Edwin Jaynes proposed to use Shannon's Entropy as the criterion to maximize, subject to all available constraints, in order to derive conservative inferences from the evidence.

In addition, if we have some non-sample prior information about the probabilities $\{p_j^0\}$, then an equivalent problem is to minimize the Kullback-Leibler directed divergence, or Cross Entropy, between the prior and the posterior probabilities (Kullback, 1959; Good, 1963). The Cross Entropy is defined as $CE = \sum_j p_j \log(p_j/p_j^0)$ if p_j^0 are the priors. Furthermore, if the prior probabilities p_j^0 are assumed to be uniform, then the Cross Entropy formalism reduces to the Maximum Entropy formalism. Not surprisingly, both the CE and the H objectives are related and really special cases of the family of Cressie Read power divergence measures (Cressie and Read, 1984). Notwithstanding the diverse types of constraints that theory may suggest (e.g., geometric moment, higher order moment, inequality constraints, etc.) and whether or not we believe their sample analogs are measured with noise, this method of using information in a sample (evidence) to recover information about social, economic, or behavioral phenomenon falls within the growing field of *Information and Entropy Econometrics*.[6]

The key requirement of this formulation is that the unknowns be proper probabilities (i.e., non-negative quantities that sum to one). This is because Shannon's entropy, as well as the Kullback-Leibler directed divergence measures, are defined in terms of proper probabilities. Zellner (1991) and Zellner and Highfield (1988) have developed this approach extensively in the econometrics field to derive a general class of distributions that satisfy various side conditions (constraints) that may be suggested/provided by economic theory.

In an important extension of their work, Ryu (1993), used this same principle to derive *regression functions* rather than *probability distributions*. Ryu (1993) showed that if the unknown quantities can be assumed to be non-negative, then the application of the Maximum Entropy (or Minimum Cross Entropy) principle can, under suitable side conditions (constraints), yields a large number of functional forms. Using the example of a production function with 2 inputs (Capital and Labor), Ryu (1993) derived the Exponential polynomial, the Cobb-Douglas, the Translog, the Generalized Cobb-Douglas, the Generalized Leontiff, the Fourier flexible form, and the Minflex-Laurent Translog production functions simply by manipulating the side conditions.

It should also be noted that utilizing the maximum entropy formalism simply with non-negative quantities—that may not be proper probabilities—is not, however, entirely new. Similar approaches are used in the field of image reconstruction. See, for example, Gull and Daniell (1978), Gull (1989), and Donoho, Johnstone, Joch, and Stern (1992) for detailed discussions.

[6]For recent theoretic and applied work in this field, see the 2002 special issue of the *Journal of Econometrics* (Vol 107, Issues 1&2), Chapter 13 of Mittelhammer, Judge and Miller (2000), the 1997 Volume (12) of *Advances in Econometrics* titled "Applying Maximum Entropy to Econometric Problems," and the Golan, Judge, and Miller (1996) monograph. See also Maasoumi (1993), Soofi (1994), and Golan (2002) for historical discussions and general surveys.

2.2.3. An information-theoretic solution to the basic problem

This brings us back to the problem at hand. The evidence we have is in the form of the constraints (2.9) and our unknowns are in the form of non-negative hazards—precisely the kind of problem for which the Maximum or Cross Entropy formalism could be applied very profitably. However, unlike Ryu (1993), where each of the unknowns are completely unrestricted (other than being non-negative), in our case, some of the hazards are just not possible. Hence, following Ryu (1993), I define a generic Cross Entropy problem but, additionally, I introduce the d_{rmn} flags into the objective function. This ensures that hazards corresponding to periods when individuals are *not* at risk of a progression will in no way influence the objective being optimized. This modified information recovery problem can be written as:

$$\min \quad CE = \sum_{rmn} d_{rmn} \left\{ s_{rmn} \log(s_{rmn}/s^0_{rmn}) \right\} \tag{2.10}$$

subject to the constraints of (2.9). Here s^0_{rmn} is an arbitrary non-negative quantity representing our prior state of knowledge. This is a constrained optimization problem (in the unknown hazards s_{rmn}) that can be solved by variational methods. The primal Lagrange function for this problem is:

$$\mathscr{L} = \sum_{rmn} d_{rmn} s_{rmn} \log(s_{rmn}/s^0_{rmn}) + \sum_k \alpha_k \left\{ \sum_{rn} x_{krn} a_{rn} - \sum_{rn} x_{krn} \sum_m z_m d_{rmn} s_{rmn} \right\} \tag{2.11}$$

where α_k are the set of K Lagrange Multipliers corresponding to the constraints (2.9). The first order conditions for this problem can be written as:

$$\frac{\partial \mathscr{L}}{\partial s_{rmn}} = d_{rmn} \log(s_{rmn}/s^0_{rmn}) + d_{rmn} - d_{rmn} z_m \sum_k x_{krn} \alpha_k = 0 \qquad \forall r, m, n, \tag{2.12}$$

so that canceling the d_{rmn} terms and solving for s_{rmn} yields the general solution:

$$s_{rmn} = s^0_{rmn} \exp\left(z_m \sum_k x_{krn} \alpha_k - 1 \right) \qquad \forall r, m, n. \tag{2.13}$$

If we assume that $s^0_{rmn} = \exp(1)$ then this yields a simple log-linear solution for the hazard. That is, we get $\log s_{rmn} = z_m \mathbf{x}'_{rn} \boldsymbol{\alpha} \; \forall r, m, n$.[7] Other assumptions are possible and will yield different solutions. More on this later.

Note that we can also use the general solution of (2.13) back in the primal *constrained* optimization problem (2.11) to derive a dual *unconstrained* optimization problem in the

[7] Here $\mathbf{x}_{rn} = (x_{1rn}, \ldots, x_{Krn})'$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)'$ are the attributes and Lagrange Multipliers written in vector notation so that $\mathbf{x}'_{rn} \boldsymbol{\alpha} = \sum_k x_{krn} \alpha_k \; \forall r, n$.

unknown Lagrange Multipliers. By inserting this solution into the primal problem we get

$$\begin{aligned}
\mathscr{L} &= \sum_{rmn} d_{rmn} s_{rmn}\Big(z_m \sum_k x_{krn}\alpha_k - 1\Big) + \sum_k \alpha_k \Big\{ \sum_{rn} x_{krn} a_{rn} - \sum_{rn} x_{krn} \sum_m d_{mrn} z_m s_{mrn} \Big\} \\
&= \sum_{krmn} \alpha_k x_{krn} z_m d_{rmn} s_{rmn} - \sum_{rmn} d_{rmn} s_{rmn} + \sum_{krn} \alpha_k x_{krn} a_{rn} - \sum_{krmn} \alpha_k x_{krn} z_m d_{rmn} s_{rmn} \\
&= \sum_{krn} \alpha_k x_{krn} a_{rn} - \sum_{rmn} d_{rmn} s_{rmn} = \mathscr{G} \qquad (2.14)
\end{aligned}$$

Given that the solution (2.13) is a function of the unknown Lagrange Multipliers α_k, (2.14) is simply an unconstrained objective function that needs to be maximized with respect to the unknown quantities α_k. That is, *minimizing* the objective (2.10) with respect to the unknowns s_{rmn} subject to the constraints (2.9) is identical to *maximizing* the dual objective (2.14) with respect to the unknowns α_k. Additionally, the dual is an *unconstrained* optimization problem therefore conventional software that contain unconstrained optimization routines (e.g., SAS, GAUSS, etc.) can be used to solve this problem.

The dual objective is a non-linear function that must be maximized with respect to the parameter vector $\boldsymbol{\alpha}$. As such, it falls under the general class of extremum estimators. The consistency and asymptotic normality of these estimators can be established under fairly general regularity conditions (Mittelhammer, Judge, and Miller, 2000, pg 132-139). However, as is evident, the objective ignores the clustering of observations within an individual. That is, individuals that have multiple arrest events are treated as contributing multiple independent pieces of information to the objective. This typically results in biased (downwards) asymptotic standard error estimates misleading us into being overly confident about our parameter estimates. To correct for this bias, following Ezell, Land, and Cohen (2003), I construct and utilize a *modified sandwich variance estimator*. Sandwich estimators (Huber, 1967; White 1980) are now very commonly utilized in econometrics and statistics when researchers are unsure about the complete specification of the distribution in a fully parametric model but are fairly sure that the mean value is well specified. The modified sandwich variance estimator merely corrects the sandwich estimator further for the possibility that there may be unobserved but persisting heterogeneity within individuals over time. Detailed analytical derivations are available from the author on request.

2.2.4. Flexible functional form and generalized hazard models

The solution described above was generic and I utilized a single set of constraints (2.9) in deriving it. However, formal theoretical reasoning and/or casual past experience may suggest many other forms of constraint each of which will alter the solution derived. For example, we may believe that attributes x_{krn} not only explain variation in the age at which particular events occur (i.e., how a_{rn} varies across n and r) but also its higher moments (e.g., how $a_{rn} \log a_{rn}$ varies across n and r). If so, then, in addition to requiring the satisfaction of constraints involving the a_{rn}, we can require satisfaction of constraints involving $a_{rn} \log a_{rn}$ as well.

To do so, we proceed in the same way as before. Let us pre-multiply both sides of the approximation (2.2) by d_{rmn}, $z_m \log z_m$, and x_{krn}, and sum over all r, m, and n. In a manner analogous to (2.9), this yields constraints of the form:

$$\sum_{rn} x_{krn} a_{rn} \log a_{rn} = \sum_{rn} x_{krn} \sum_{m} z_m \log z_m d_{rmn} s_{rmn} \qquad \forall k \in K. \tag{2.15}$$

This set of constraints does not need to have the same attributes as (2.9). I am assuming that they are the same for ease of notation. Now, our information recovery task can be modified to a constrained optimization problem subject to the two sets of constraints *simultaneously*. Following the same derivations as above, we can derive the optimal solution as:

$$s_{rmn} = s^0_{rmn} \exp(z_m \mathbf{x}'_{rn} \boldsymbol{\alpha} + z_m \log z_m \mathbf{x}'_{rn} \boldsymbol{\beta} - 1) \qquad \forall r, m, n, \tag{2.16}$$

where, $\boldsymbol{\beta} = \beta_1, \ldots, \beta_K$ are a new set of Lagrange Multipliers corresponding to the constraints (2.15). Moreover, as in the simpler case, we can convert the constrained minimization problem into an unconstrained maximization problem in the unknown Lagrange Multipliers (both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) simultaneously. This dual objective takes the form:

$$\mathscr{G} = \sum_{krn} \alpha_k x_{krn} a_{rn} + \sum_{krn} \beta_k x_{krn} a_{rn} \log a_{rn} - \sum_{rmn} d_{rmn} s_{rmn} \tag{2.17}$$

where s_{rmn} is the solution given in (2.16).

An appropriate definition of d_{rmn} can help restrict the analysis to a single event, to include censored cases, and/or to remove individuals from the risk pool. For censored cases, as noted in the non-parametric derivation, we simply define $d_{rmn} = 1$ when the individual is at risk of experiencing the next event until (s)he is censored. After that, we set $d_{rmn} = 0$. This means that, on the LHS of the constraints, we will have the actual values of a_{rn} and/or $a_{rn} \log a_{rn}$ only when the event is observed but a value of 0 when the case is censored. To see this note that the term $\sum_m d_{rmn} y_{rmn} z_m = 0$ for all censored cases because $y_{rmn} = 0 \; \forall m$ when the record is censored. Hence, in addition to redefining the set of dummy variables d_{rmn} appropriately, if we let $c_{rn} = 1$ flag the censored cases, and re-define the age variables as

$$a^*_{rn} = \begin{cases} a_{rn} & \forall c_{rn} = 0 \\ 0 & \forall c_{rn} = 1 \end{cases} \quad \text{and} \quad a^*_{rn} \log a^*_{rn} = \begin{cases} a_{rn} \log a_{rn} & \forall c_{rn} = 0 \\ 0 & \forall c_{rn} = 1 \end{cases} \tag{2.18}$$

then we can use a^*_{rn} and $a^*_{rn} \log a^*_{rn}$ in the objective function (2.17) when the data include censored cases. The remaining derivations remain unaltered.

More generality can, of course, be built into this framework by assuming a general set of constraints that involve various transformations of a_{rn}. These could include linear, quadratic, cubic, quartic, Fourier, log-linear, etc. By adjusting these constraints, we can derive a large number of parametric forms. Moreover, we can impose sets of these constraints *simultaneously* to get more generalized forms that nest several alternatives and test for specific functional forms.

The framework presented here is not, however, entirely innovative. In a recent survey of dynamic duration models, Ebrahimi and Soofi (2003) show how several of the standard parametric models along with several mixture models by utilizing an information-theoretic objective while specifying differential equation constraints that govern the evolution of the hazard over time (See also Table 1 in Soofi, Ebrahimi, and Habibullah [1995]). Other recent articles involving the same principle—what the authors refer to as the principle of *Minimum Dynamic Discrimination* or *Maximum Dynamic Entropy*—include Ebrahimi, Habibullah, and Soofi (1992), Ebrahimi and Kirmani (1996), and Asadi, Ebrahimi, Hamedani, and Soofi (2005).

The framework I present in this report builds on this literature but utilizes a discrete support thereby negating the need for differential equation constraints and, following Ryu (1993), I formulate the objective in terms of the hazard directly (rather than the underlying probability distributions). This adds considerable computational efficiency.

2.3. DEFLECTING OFFENDING TRAJECTORIES

2.3.1. Estimating the deviation of trajectories from counterfactual paths

So far we have not made any explicit assumptions about the priors s^0_{rmn} except noting that if we fix it to exp(1), we obtain a simple log-linear specification for the path. If we do have some prior knowledge about the evolution of the hazard over time, we can introduce that information in the form of the s^0_{rmn} so that the final solution is computed as a *deviation* from this prior. This formulation is particularly relevant for our analysis since we wish to study the deviation of a trajectory from a counterfactual. But first, we need to construct a plausible counterfactual.

A simple way to construct this counterfactual is to model the links between age, arrest number, and other attributes using the framework described above but by estimating it only with the pre-prison part of the available arrest histories. This model would therefore capture the dynamic process by which individuals in the sample had been accumulating their arrest histories prior to prison admission. Next, using this model, we can project a future trajectory (from the age at release onwards) using knowledge about the arrest number this particular individual was at risk of as well as all the other attributes as $\tilde{s}_{rmn}$. These projections trace out the entire evolution of the hazard *for the next arrest* over the remaining life of the individual given knowledge about the past criminal history accumulation process. As such, each provides the perfect counterfactual for assessing future offending patterns since this is the path we should expect the releasee to have been on at the time of release *had (s)he not been incarcerated*. Therefore, when we model the post-release offending trajectory—i.e., the hazard of the next event in the sequence of arrests—we simply replace s^0_{rmn} with $\tilde{s}_{rmn}$ in the dual objective function (2.17). This yields a solution exactly like (2.16) where $s^0_{rmn} = \tilde{s}_{rmn}$.

Why would this procedure model the deviation from the pre-prison based counterfactual trajectory? To see why, consider the case where all parameters in this post-release model are found to be 0 (i.e., $\alpha_k = 0$ and $\beta_k = 0\ \forall k$). We then obtain the result that $s_{rmn} \equiv \tilde{s}_{rmn}$. In other words, if all the parameters of the post-release model are zero then there has been no deviation from the path the individual was projected to be on—i.e., the counterfactual. To the extent that these parameters are non-zero, there has been a deflection of the trajectory as a result of this incarceration experience. What remains then is to find a way to decide whether this deflection is for the better (lowered trajectory compared to the counterfactual), worse (higher trajectory compared to the counterfactual) or about the same. I derive one such measure next.

2.3.2. Classifying the incarceration experience

Ebrahimi and Soofi (2003), present a method for comparing information across two hazard paths (either across individuals or across two different paths for the same individual) that is particularly well suited for comparing the evolution of two trajectories over time. Their approach utilizes the notion that the Kullback-Leibler directed divergence measure (or Cross Entropy) is a measure of divergence between two probability distributions. Since probability distributions and hazards are two different ways of representing the same underlying phenomenon, they derive dynamic divergence measures between the evolution of two hazard functions.

Applying this idea in our case is fairly straightforward. Since the objective is defined in terms of the natural log of the ratio of two strictly positive numbers, then

$$\log(s_{rmn}/s^0_{rmn}) \quad \begin{cases} > 0 & \text{iff } s_{rmn} > s^0_{rmn} \\ = 0 & \text{iff } s_{rmn} = s^0_{rmn} \\ < 0 & \text{iff } s_{rmn} < s^0_{rmn} \end{cases} \qquad \forall r, m, n. \tag{2.19}$$

The problem with this measure, as it stands, is that it is a function of age and therefore it can, and typically will, be different for each m. What we need is a way to aggregate across this divergence measure over the entire *residual life* starting from any point z^*_m (e.g., the date of release).

Ebrahimi and Soofi (2003) present a way to approach this problem by redefining the hazards into probabilities and noting that the measure reduces to the traditional Kullback-Leibler divergence measure with an appropriate normalization and a ratio of survival functions (Ebrahimi and Soofi, 2003, pg.6). In an analogous, but unrelated derivation, Ryu (1993) showed that the Maximum Entropy solution for any positive quantity could be considered an *averaged density* if we normalize appropriately. In our case, the quantity of

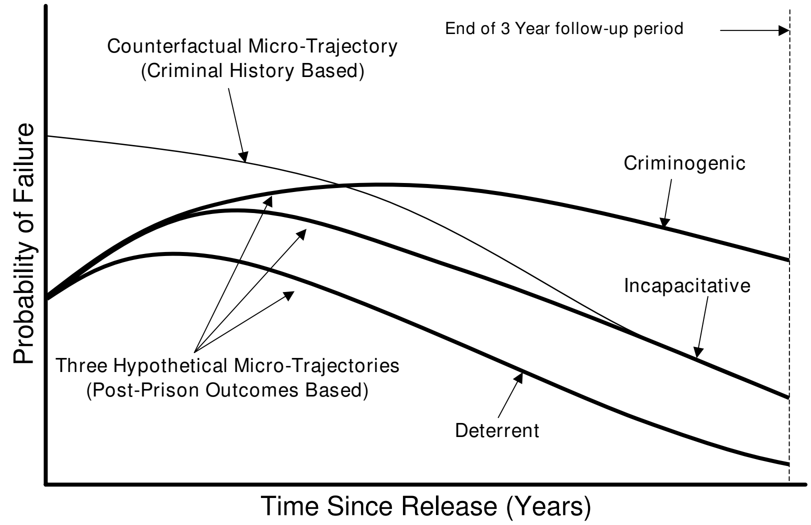


Figure 2.2: A counterfactual trajactory compared with three hypothetical post-release trajectories showing criminogenic, incapacitative, and deterrent effects of incarceration.

interest is the hazards for all points beyond the date of release. Hence, following Ryu (1993), if we define the term $s^*_{rmn} = \sum_m d_{rmn} s_{rmn}$ for some appropriately defined d_{rmn}, then we see that

$$\pi_{rmn} = \frac{d_{rmn} s_{rmn}}{s^*_{rmn}} = \frac{d_{rmn} s_{rmn}}{\sum_m d_{rmn} s_{rmn}} \tag{2.20}$$

is a proper probability wherever it is defined (i.e., $\sum_m \pi_{rmn} = 1 \; \forall r, n$ and $\pi_{rmn} \geq 0 \; \forall r, m, n$). This means that the objective function we are optimizing already contains information about the averaged difference between $\tilde{s}_{rmn}$ and s_{rmn}. All we need to do is normalize the objective appropriately. This normalization provides a way to aggregate the various terms in the trajectory (2.19) across the entire residual life of the individual upon release. This measure is defined as:

$$\begin{aligned} \delta_{rn} &= \frac{1}{s^*_{rmn}} \sum_m d_{rmn} s_{rmn} \log(s_{rmn}/s^0_{rmn}) \\ &= \sum_m \frac{d_{rmn} s_{rmn}}{s^*_{rmn}} \log\log(s_{rmn}/s^0_{rmn}) \\ &= \sum_m \pi_{rmn} \log(s_{rmn}/s^0_{rmn}) \end{aligned} \tag{2.21}$$

The δ statistic can be seen as an average (expected) log divergence between the projected trajectory (based on knowledge about pre-prison arrest patterns) and the actual post-release offending pattern. An example of a counterfactual and three hypothetical post-release trajectories are shown in Figure 2.2. As shown there, the trajectories can be different at any given point in the post-release period. However, the δ statistic derived above measures divergence between two *paths* rather than points. Moreover, since π_{rmn} is a proper probability (summing to one), we can compute the standard deviation of this quantity (the log divergence) as well. The standard deviation of each δ_{rn} can be computed as:

$$\sigma_{rn} = \sqrt{\sum_m \pi_{rmn}\Big(\log(s_{rmn}/s^0_{rmn})\Big)^2 - \Big(\sum_m \pi_{rmn}\log(s_{rmn}/s^0_{rmn})\Big)^2} \qquad \forall r, n, \qquad (2.22)$$

which follows from the definition of the variance of a random variable v as $E(v^2) - E(v)^2$.

Finally, we can utilize these definition of δ_{rn} and σ_{rn} to decide whether the expected log divergence of the residual life trajectories are *sufficiently* different. The current incarceration is deemed to have had an

$$\begin{array}{rll} \text{Deterrent Effect} & \text{iff} & 0 < \delta_{rn} - 2 * \sigma_{rn} \\ \text{Incapacitative Effect} & \text{iff} & 0 \in \delta_{rn} \pm 2 * \sigma_{rn} \\ \text{Criminogenic Effect} & \text{iff} & 0 > \delta_{rn} + 2 * \sigma_{rn} \end{array} \qquad (2.23)$$

These classifications allow one to model the effects of individual, contextual, and policy options on the likelihood of a releasee's prison experience being one of the three types. This can be done in standard software using multinomial discrete choice models or ordered discrete choice models. Such an analysis could be used, for example, to study what measures can increase the likelihood of the deterrent experience and minimize the likelihood of a criminogenic experience.

Alternately, one can study the effects of these individual, contextual, and policy options on the continuous variable δ_{rn} directly since larger '+' values of δ indicate larger criminogenic effects and larger '−' values of δ indicate large deterrent effects. In the next chapter, I explore both these approaches and present a limited set of results.

2.4. CONCLUDING COMMENTS

In this chapter, I developed an information-theoretic framework for modeling the detailed criminal history accumulation process of a group of releasees. There exists, of course, several other methods that are capable of modeling event histories (see, for example, Mayer and Tuma [1990] and Blossfeld, Hamerele, and Mayer [1989]). The method developed here has several benefits over existing strategies.

First, unlike fully parametric functional forms, the information-theoretic approach allows an easy incorporation of several constraints that yield flexible functional-form hazard models. Under restrictive assumptions, this approach yields several of the standard hazard models as special cases. As such, the approach can be used to develop models that nest several parametric forms as special cases in order to test (statistically) assumptions about the shape of the evolution of the hazard over time or assumptions about proportionality.

Second, given its particular emphasis on minimizing the directed divergence between a prior and posterior trajectory, the approach offers an easy method for assessing whether or not the evolution of the hazard over the residual life (defined at any appropriate point, e.g., the date of release) is different from a counterfactual. The average log divergence between the two trajectories provides a convenient summary statistic for this purpose. Moreover, this statistic is not an ad-hoc measure. It is merely a re-normalized version of the *very* objective that is being optimized to obtain the hazard models.

Finally, this average divergence measure can then simply be converted into a classification or can be viewed as a continuous measure. Large negative values on this statistic imply large deterrent effects whereas large positive values on this statistic imply large criminogenic effects. Studying how this measure correlates with various attributes as well as policy options can be of immense use to practitioners and policy makers in understanding the factors that may maximize deterrent benefits of incarceration and/or that minimize the criminogenic harm resulting from it. These factors can include not only demographic factors that are outside the control of policy makers but also factors like participation in prison programs, post-release supervision/assistance programs, as well as socio-economic, behavioral, and contextual factors such as the availability of employment opportunities, family bonds, and individuals' mental health.

The method developed here is designed to take full advantage of dated criminal history records when such information *is* available. In ignoring this information, when it is available, researchers risk wasting valuable information and thereby forgo learning opportunities. Aggregate measures of criminal history scores typically use only one source of variation in the pre-release arrest history—the number of prior arrests (weighted or unweighted). The method presented here utilizes another source of variation available in the pre-prison arrest history—the process by which this arrest history was accumulated. Furthermore, it utilizes this knowledge in informing the future evolution of the hazard.

To be sure, the method described here is not the only way one can study trajectories of offending patterns over time. There exists a large literature in criminology that aims to model the trajectories of offending patterns over the life course of individuals using group based modeling techniques (Nagin, 2005; Nagin and Land, 1993; Land and Nagin, 1996). Responding to concerns raised by Hagan and Palloni (1988), in particular, Land and Nagin (1996) demonstrated that group-based trajectory models are well suited to take into account the order of arrest events. Similarly, the approach developed here is not incompatible with approximating unobserved heterogeneity via finite mixture modelling strategies. Therefore, it would be a profitable extension of the current work to include distinct group-based heterogeneity in the models as well. However, for the approach to have practical utility, the emphasis should remain on attempting to construct counterfactual trajectories for each and every individual in the sample (not just for groups). In this report, I rely solely on available attributes to model the heterogeneity in the evolution of the hazards.

Chapter 3

An Application

In this chapter, I apply the methods developed in the previous chapter to a specific data set. The chapter is organized as follows. I begin with a brief description of the dataset and the variables used in this report. I then summarize estimates of the pre-prison based criminal history accumulation process and discuss the findings. I also use these models to make projections for individuals at the time of their release. These projections are compared with the actual arrest events post-release. Next, I use these projected counterfactual trajectories as a backdrop against which to develop the actual post-release offending trajectories. These model estimates are also presented in summary form. Finally, using the methods developed in the previous chapter, I compute the δ statistic and use it to classify individuals' incarceration experiences. I present a limited set of results from standard linear and logistic regressions that are used to model variation in these experiences across individuals.

3.1. THE DATA

3.1.1. Data source

The data used in this research effort is available to the public from the National Archives of Criminal Justice Data (NACJD), at the Inter-University Consortium for Political and Social Research (ICPSR), University of Michgan, Ann Arbor, MI. It is archived as study # 3355 (*Recidivism of Prisoners Released in 1994 [United States]*) (BJS, [2002]).

The data were collected by the Bureau of Justice Statistics (BJS). It tracks a sample of 38,624 prisoners released from 15 state prisons in 1994 over a period of 3 years. The vast majority of the archived database consists of information on each releasee's entire officially recorded criminal history. This includes all recorded adult arrests through the end of the follow-up period. These data were obtained from state and FBI automated RAP sheets which include arrest, adjudication, and sentencing information. Each arrest event includes information on adjudication and sentencing related to that event if such action was taken. Unfortunately, however, the data do not contain detailed information on when these individuals were released from prison if they were imprisoned after a particular arrest event.[1]

In addition to the detailed dated event history data, this database also contains a limited amount of demographic and related information. Demographic measures available in the database include date of birth, race, ethnicity, and gender. Some detail is available about the type of release from prison (e.g., parole, mandatory release, etc.) and some about the type of admission into prison (e.g., new court commitment, new court commitment with a violation of conditions of release, etc.). However, this information is available only for the 1994 release and not for all prior (or future) arrest events.

Since the emphasis of this effort was to develop an analytical approach, I have restricted the analysis to rearrest only. Application to other type of outcomes (reconvictions, reincarcerations, self-reported offending patterns, or relapse into drug involvement, etc.) is straightforward.

Before conducting the analysis, some diagnostic checks were run on the data to ensure they were compatible with the model requirements. Since the data are based on official records and possible disparate sources of date information (e.g., date of birth obtained from the state data and from the FBI data could differ), I first computed the ages for each of the arrests in the data. Then, I checked for the chronology of these dates and checked to see if the age variable was well defined. I created flags for any *individual* that had records that were not in proper chronological order or whose ages were incorrect/impossible (e.g., negative or below 15). In addition, I created flags that identified any individuals that were missing information on all ages or that had gaps in their age variable. For example, individuals that had appropriate ages for the first and second arrest events but were missing age on the third event and again had appropriate ages for all subsequent arrests were flagged as potentially problematic. After creating these flags, I performed a list wise deletion of records—i.e., all records for individuals with any problem (as determined by the various flags) were dropped from the analysis set.

Additionally, the data contains a variable ANALYSIS that flags all records that were included in the BJS report titled *Recidivism of Prisoners Released in 1994* (Langan and Levin, 2002). The criteria for inclusion in the report are provided on page 14 of Langan and Levin (2002). In my analysis, I also excluded all persons that were not included in BJS's report (i.e., persons flagged as ANALYSIS=0).

3.1.2. Data structure

After removing persons who either had some problem in their arrest histories or were not included in the BJS report, the remaining sample consisted of 32,628 persons across 15 states. In addition, since the sample for California releasees was very large (nearly 60,000

[1]This implies that the data are unable to calculate street time. However, the data do provide information on the adjudication outcome at each successive arrest events that I utilize in the models.

person-events before prison release) I used a random subset of 2500 individuals (21,838 person events) from the California sample for estimating the pre-prison criminal history accumulation process. For the analysis of the post-release data, however, all individuals from California were included in the study. The final pre-release dataset therefore consisted of 21,226 individuals across the 15 states whereas the post-release data consisted of the 32,628 individuals.

Arrest records for these persons were next re-structured into a hierarchical person-event level file. That is, arrest events of each person were all clustered in chronological order. The arrest histories were next truncated after the first post-release re-arrest event. Recall that, for the post release period, we are only examining the first rearrest event. For those persons that were not arrested after release, the arrest age was set to the age at censoring (i.e., release age + 3 years).

The data were structured similar to the arrest profiles displayed in Table 2.1. In addition to the key criterion variable—age at arrest—the data were also manipulated to create a set of individual level fixed covariates as well as covariates changing over time.

3.1.3. Key variables included

The key independent variables used in estimating the pre-release criminal history accumulation process included the arrest number (EVENTNUM), the age at first arrest (AGE1ST), whether or not the individual was confined as a result of the previous arrest event (CONFLAST), and a measure of the number of years taken to reach each arrest event cumulated through the last arrest event (CARAGE). AGE1ST and CONFLAST were set to 0 for the first arrest event.

Besides CARAGE, the variables used in this part of the analysis are self explanatory. CARAGE was defined as a measure that captures the evolution of the heterogeneity in the sample members as they aged. It is defined as

$$\text{CARAGE}_{rn} = \sum_{j=1}^{r} \frac{a_{jn}}{j} \qquad \forall r, n, \tag{3.1}$$

and it captures variation in the past criminal history up to the current arrest in such a way that it distinguishes people who are closer to their past arrest "clusters" from those that are further. Table 3.1 shows hypothetical past arrest histories of two individuals and demonstrates the calculation of CARAGE at each arrest event. Note that both individuals have the same CARAGE until their 2nd arrest because they follow the same path. As they differ in their arrest patterns CARAGE begins to record this heterogeneity. In fact, individual A gets a higher CARAGE on his 3rd arrest because he is "closer" to his past arrest cluster at age 30 than individual B is at age 35. After that, both individuals are rearrested at age 40 but their CARAGE continues to record their heterogeneous pasts. In this sense, the variable records heterogeneity in past offending patterns and, all else being equal, assigns a higher

to supervision, and unconditional release. For others (AZ, FL, NJ, and OH) the only available information was whether the release was CONDITIONAL or otherwise. Finally, for MN and OR the only available information was whether the release was for PAROLE or MANDATORY release. Hence, when analyzing the effects of release type on the likelihood of the prisoner's experience being deterrent or otherwise, separate models were estimated for groups of states to increase statistical power.

VIOLENT, PROPERTY, and DRUG refer to the most serious offense for which the prisoner was serving time when (s)he was released in 1994.

Finally, note that for some states the average age at which persons recidivated (RECIDAGE) would seem to be at or below the average age at which prisoners were released. However, this is misleading because the age of recidivism is computed *only* for those that were rearrested within the follow-up period. Similarly, the age at censoring (CENSORAGE) is computed *only* for those that were censored within three years of release.

Before, proceeding with the estimation and analysis of the hazard models, I first conducted some simple graphical diagnostics. I present those next.

3.2. PREDICTABLE PATTERNS

Before proceeding with model estimation, it would be good to assess whether the arrest histories contain any predictable patterns. After all, the entire strategy rests on such patterns existing. Moreover, this release cohort is a mixture of several birth cohorts and one might consider the sample too heterogeneous to capture in a single model. To that end, I first construct some basic Kernel density plots of the ages at various arrest events. The density plots for arrest events 1 though 20 (DEN01 - DEN20) are presented in Figure 3.1.

As is evident, there is a very predictable pattern visible. The pattern has two components. First, the age distribution of each successive arrest shifts slightly to the right as we go from lower arrest numbers to higher. Second, the dispersion of the distribution increases as we move from lower to higher arrest numbers. Recall that our flexible hazard model utilizes precisely these moments to recover information about the trajectories. Hence, if we are able to capture the process underlying these distributions, we should be able to project with fair amount of confidence what we could have expected in the absence of incarceration.

Of course, we cannot simply model the age distribution directly because this masks the dependence structure between successive arrest events. Therefore, we need to be able to model the hazards appropriately while using the predictable pattern observed in Figure 3.1. The formulation of the flexible functional form models in the previous chapter afford us that opportunity. Note, for example, that the constraints that we impose in (2.9) and (2.15) explicitly link the unknown hazards to the first two moments of the age of arrest. One of the reasons we typically need to model the second moment is if there is reason to believe that the pattern has systematic over or under-dispersion. If not, a simple moment based model (e.g., Poisson) of the age distribution would suffice. Figure 3.2 plots the mean age

and its variance for these 20 arrest events. Again, it is evident that the variance exceeds the mean at all arrest events *and* that the variance evolves in a non-linear way—first reverting towards the mean rapidly and then moving towards it more slowly.

3.3. MODELS OF THE CRIMINAL HISTORY ACCUMULATION PROCESS

In this section, I present the model estimates of the pre-release criminal history accumulation process. In order to keep the estimation manageable and to afford the model full flexibility, I estimated separate models for each of the states. The form of the model is held fixed across all state samples.

First, I present some evidence that the clustering of observations does in fact provide for biased (typically downward) standard errors. Consider, the model for Arizona. Table 3.3 shows the results of the information-theoretic model and presents three sets of asymptotic standard errors. The first set (TRAD) are those computed by inverting the negative Hessian of the dual objective function, the second set (SAND) are the sandwich estimates, and the third set (MODS) are the modified sandwich estimates. As is expected, the sandwich estimates of the standard errors are higher than the traditional estimates and the modified sandwich estimates are higher still. This is because both the traditional as well as the simple sandwich estimates ignore the clustering of the observations. Although, in this example, all the parameters remained statistically significant irrespective of the a.s.e. estimate used, as is evident from the various Wald-χ^2 values provided, there are huge reductions in the confidence we have about several of these Lagrange Multipliers when we account for the clustering. For the rest of this report, therefore, I err on the side of caution and use the modified sandwich estimates for making inferences.

Since the models are formulated in terms of hazards, a negative Lagrange Multiplier implies that the variable in question decreases the hazard's path or, put another way, the variable in question increases the expected duration to the next event. As such, the negative values of the parameters for EVENTNUM are consistent with Figure 3.1 and Figure 3.2. That is, increases in arrest numbers are associated with higher age (duration from birth to event). Moreover, the *positive* sign on the corresponding β_k multipliers suggests that the increasing age associated with increasing event numbers is at a decreasing rate. This simply means that the relationship between the arrest number and the hazard trajectory is non-linear. Note that all the β_k parameters have the reverse sign relative to the corresponding α_k parameters.

Similarly, increases in age at first arrest are associated (as expected) with increasing age at subsequent arrest (i.e., decreasing hazard paths for subsequent events). Moreover, this relationship is non-linear. CARAGE, also as expected, has a positive coefficient in the hazard model. Recall that CARAGE measures the closeness to past clusters of arrests. As such, a positive coefficient in the hazard model suggests that being close to a prior cluster decreases the duration and increases the hazard of the next event. As with the other parameters, this too has a non-linear link with the outcome of interest.

Consider, for example, a situation where all parameters are found to be statistically indistinguishable from 0 (i.e., insignificant). That would mean that the post-release trajectory is statistically indistinguishable from the prior (i.e., the counterfactual). Hence, if one or more of the parameters are found to be significantly different from 0, this would indicate that, in the sample as a whole, there has been a deviation of at least *some* of the post-release trajectories from their counterfactuals. It should not be taken to mean that *every* trajectory has deviated from its counterfactual.

I present the results of the post-release sample in Table 3.5 in a manner analogous to the presentation in Table 3.4. The pattern of coefficients are different from those in Table 3.4. That is to be expected. However, unlike the pre-release models, the post-release model parameters vary a lot more across states. Moreover, some parameters even take the opposite signs. For example, the value of α_k for AGE1ST is positive and significant for DE but is negative and significant for IL. In a similar manner, the signs of the significant values of β_k for AGE1ST vary considerably across states. This suggests that the way trajectories are deflected between the pre- and the post-release periods varies across states and that the effects of AGE1ST in particular can even be reversed across different states.

On the other hand, there are some factors that exert unambiguous pressure on offending trajectories. Being later in the criminal career exerts an upward pressure on the offending trajectory relative to the counterfactual. That is, large values of EVENTNUM are associated with an upward pressure on the offending trajectory. Similarly, being closer to past cluster exerts a downward pressure on the trajectory relative to the counterfactual. As noted above, these are aggregate statements about the sample as a whole. The actual deflection for each and every releasee will be computed in the next section and the determinants of these *individual-level* deflections will be investigated there.

Signs of the deflection of the trajectories can be seen directly in the projected rearrest rates as well as the false positive and false negative rates. Although the prediction problem is no longer an out-of-sample one, simple comparisons between these projected rearrest rates and the counterfactual projections of the last section shows that the post-release projections are far superior to those of the counterfactuals. With the exception of MI and OH, where the false positive rates are about 40%, we see that the false positive rate typically is between 25-30%.

As further evidence that the posterior trajectories are sufficiently deflected from the priors, we can compute and plot the short-term (quarterly) predictions from these models. Figure 3.6 shows these curves. Cursory comparison with Figure 3.5 shows that the false positive rate is substantially lower. Moreover, within about 6 months of release, both the false positive and false negative rates are below 50% and then remain stable.

merely provided to demonstrate the approach. I do not make any recommendations based on these results. Much more detailed information is be needed before explicit policy recommendations could be enumerated.

3.5.1. Discrete choice models of the incarceration experience

First, consider simple logistic regression models of whether or not a releasee's prison experience will be deterrent (rather than merely incapacitative or even criminogenic). Since the proportion of releasees that were deemed to have had a criminogenic experience is fairly small, I combine these classified as having had a criminogenic and incapacitative experience into one category. Hence, the estimates I present in Table 3.6 are from models that attempts to link various available attributes to the likelihood of being deterred versus not. Once again, the table presented here only summarizes the signs of the various predictors in affecting the likelihood a deterrent effect. Detailed coefficient estimates are provided in the Appendix.

Four sets of parameters are presented there. One of the key policy variables to be investigated —the type of release from prison—was not consistently available in all states. The variable was re-coded into discretionary release to supervision (PAROLE), mandatory release to supervision (MANDATORY), and unconditional release (UNCONDITIONAL). Based on this variable, I collapsed states into 4 groups. Group I includes all states that had sufficient detail to model the effects of various types of release mechanisms (MD, NY, NC, TX, and VA), Group II included states that only allowed a comparison of discretionary release to mandatory release (MN and OR), Group III included states that only allowed a comparison of conditional versus unconditional releases (AZ, FL, NJ, and OH), and Group IV included states that did not contain enough variation to permit estimating the effects of this policy variable on the effects of incarceration (CA, DE, IL, and MI).

Since the logistic regression models were predicting the probability of deterrent experience, therefore positive and significant coefficients can be expected to increase the likelihood of a releasee having been deterred as a result of this incarceration. Similarly, negative coefficients imply increased likelihood that the releasee had merely an incapacitative or even a criminogenic experience.

As should be expected, releasees that have higher numbers of prior arrests are *less* likely to experience deterrent effects. Those closer to their prior arrest clusters and those released later in life were *more* likely to experience deterrent effects. Surprisingly, those with later ages of first arrest were consistently less likely to experience deterrent effects. Among the Group I states, Blacks were more likely to experience deterrent effects while among Group IV states they were less like to be deterred. Males were less likely to be deterred by incarceration (among the states in Groups I, II, and IV) and, typically, prisoners released from Violent, Property, or Drug related crimes were less likely to experience deterrent effects (relative to Public Order crimes). Surprisingly, the release mechanism seemed to have minimal effect in explaining the type of experience release could expect. The only

the classifications.

On the other hand, the linear regression findings lend strength to some of the other conclusions reached at using the logistic regression analysis. For example, RELAGE and AGE1ST have the same signs across all state groups *and* they have the same qualitative effect on the incarceration experience as was found in the categorical analysis. In a similar manner, females are (in this model) unambiguously more deterred by their incarceration experience than males. The offense for which releasees were incarcerated seems to have little or no contribution towards explaining variation in the deterrent effects. Finally, discretionary release (when compared with mandatory release) has a higher deterrent effect on releasees and mandatory release (as compared to an unconditional release) seems to have a higher deterrent effect on releasees. Similar sporadic findings of a deterrent effect of release mechanism were also found in the logistic regressions.

Hence, when used in concert, the two sets of analysis have the potential of strengthening the conclusions one may reach about the kinds of factors that can be expected to increase the deterrent benefits of incarceration and minimize its criminogenic harm.

3.6. SUMMARY OF FINDINGS

The research conducted, and reported on here, was largely a development effort. Despite that, some interesting findings emerged from the effort that are summarized below.

1. There was a fair amount of consistency among all the pre-prison based models of the criminal history accumulation processes across the 15 states analyzed. For example, being further along in the criminal career (i.e., being at risk of a higher arrest number) and starting the career later (i.e., having a higher age at first arrest) pretty consistently result in lowered hazard trajectories. Similarly, all else being equal, being closer to past arrest clusters, is consistently associated with increased hazard trajectories. There was less consistency among states when modeling the deviation between the counterfactual and actual rearrest hazard trajectories after release. Being later in the criminal career exerts an upward pressure on the offending trajectory relative to the counterfactual. Similarly, being closer to past cluster exerts a downward pressure on the trajectory relative to the counterfactual.

2. The criminal history accumulation process contains valuable information about the long-term secular trends in individuals' offending patterns over the life course. The counterfactual trajectories, based on estimated models of the pre-prison based criminal history accumulation process and projected for the post-release period, perform remarkably well in predicting rearrests within three years of release.

3. As expected, the same counterfactuals do not perform as well when used for making short-term projections. The false-positive rates are at very high levels throughout

the follow-up period. When updated with models of the post-release behavior, the models perform remarkably well.

4. In this analysis, large portions of the release cohort were classified as having had an incapacitative or a deterrent experience. A small proportion of the sample experienced criminogenic effects as a result of this incarceration.

5. Using these classifications as the criterion outcome, increased age at release and being closer to past clusters were consistently found to *increase* the likelihood of a releasee experiencing a deterrent effect. Having more prior accumulated arrests and having a later age at first arrest were both found to significantly *decrease* the likelihood of a deterrent effect. Being released to supervision was found not to deter releasees substantially.

6. Using the average log divergence between the counterfactual and the actual trajectories as the criterion some anomalous findings were uncovered. However, the effects of age at first arrest and age at release were qualitatively similar to what was found in the categorical analysis. Additionally, females were expected to experience larger deterrent effects than similar males.

3.7. CONCLUDING COMMENTS

In this chapter, I applied the analytical framework developed in this research effort to a particular data set. I estimated several models of the pre-prison criminal history accumulation process and used that to construct counterfactual trajectories for future offending patterns. Models of post-release offending trajectories for the next rearrest event were then estimated using the projected counterfactuals as prior knowledge. Furthermore, the post-release trajectories were compared with the counterfactuals for each individual in the post-release sample and values of the δ statistic (the average log deviation of the actual and counterfactual) were computed. Using the expected value of δ and its standard deviation, the current incarceration experience of each of the sample members was classified as having been deterrent, criminogenic, or merely incapacitative. Finally, simple models linking these experiences to available attributes were estimated and discussed.

The point of this exercise was to demonstrate the capabilities of the approach. As noted in the introductory chapter, no specific policy recommendations can or are being made as a result of this analysis. Its sole purpose was to develop and explain the analytical framework.

Before concluding this chapter, a point of clarification is in order. In this chapter, I have modeled the post-release trajectory as the evolution of the hazard *for the next arrest event* upon release. It is possible, though not explored in this report, to model the evolution of hazards *for all future arrest events* after release. It is also possible to compute counterfactuals for each of these future rearrest events using estimates from the pre-prison based models. However, it seems somewhat awkward to speak of the deterrent, incapacitative and criminogenic effects of the current incarceration on the 2nd, 3rd, and subsequent rearrest after release. Consider, for example, if we are to find that, for a particular individual, this incarceration had an incapacitative effect for the 1st rearrest after release but a criminogenic effect for the 2nd rearrest after release. What are we to make of this finding? It seems to me cleaner to restrict these classifications to just the first rearrest event. Alternately, one could computed composite δ measures aggregated not only across the entire residual life of the release but across all subsequent rearrest events or aggregated for specific time periods (e.g., first six months following release). Exploring these extension are promising areas for future work.

Chapter 4

Conclusion

In this chapter, I discuss the larger implications of this research effort and propose some promising directions for future research.

4.1. IMPLICATIONS

The analytical framework developed as a result of this research effort has important substantive, methodological, and practical implications.

4.1.1. Substantive implications

Substantively, the analytical framework developed here has the potential to shed light on a very important question: *How does incarceration affect individuals?* Although theoretical arguments for and against the use of incarceration as a crime control strategy abound, it is hard, in my opinion, to imagine that this important policy tool has the same effect on all persons at all ages and all times. As such, it would be very beneficial to be able to determine, or at the very least investigate, the types of individuals likely to be deterred by incarceration. In a similar way, it would be very beneficial to understand how incarceration can have differing impacts on the same people at various stages in their life and/or criminal careers. The framework developed here offers one way to directly investigate these issues.

There are several related substantive benefits that can be derived by extending this research in appropriate directions that are discussed in more detail below.

4.1.2. Methodological implications

When the detailed dated arrest histories of a sample of releasees is available to researchers, utilizing only one source of variation in the data—the total amount of criminal history accumulated prior to prison admission—for modeling the risk of future recidivism forces analysts to waste valuable information and thereby forgo learning opportunities. A second source of variation available in these pre-prison arrest histories—the process by which individuals were accumulating these histories—contains immense amount of information about future offending patterns. The information-theoretic event-history models, developed in this research effort, allow this knowledge to be introduced into the modeling strategy in a very intuitive and easy way. The process by which individuals accumulate their pre-prison arrest histories, typically, have some very predictable patterns that can be modeled. These models allow simple projection of the risk (hazard) of future arrest events. These projections can be thought of as person specific micro-trajectories that trace out the evolution of rearrest hazards *had the individual not been incarcerated*. As such, they are perfect counterfactuals against which to assess the post-release offending patters.

Statistical concepts such as Kullback-Leibler Directed Divergence measures, the family of Cressie-Read Power Divergence measures, and Information Entropy, are all inequality measures that capture the divergence between two probability distributions. Modifying these measures to capture divergence between two functions is straight forward. Therefore, building on information-theoretic foundations, divergence measures between counterfactual and actual trajectories can be developed that allow for a systematic definition of what it means for two trajectories to diverge *sufficiently*. One such measure was developed in this research effort. These measures allow for a simple classification of releasees into groups that were either deterred by their incarceration or were merely incapacitated or, in fact, had a criminogenic experience. Traditional modeling approaches, such as logistic and linear regressions, can then be used to investigate the correlates of these experiences.

Flexible functional form models of recidivism offer the possibility of increasing the predictive accuracy of the model because they are not bound by the assumptions of a particular functional form. For example, if researchers are unsure about the proportionality assumption, they may simultaneously impose proportionality and non-proportionality constraints using the data. That way, to the extent that one or the other models satisfies the real process generating the data, relevant Lagrange Multipliers will be distinguishable from 0 and the remaining will not. These flexible functional form models rely on, what is termed, the *Encompassing Principle* (see Chapter 14 in Hendry [1995]). They offer a nice way to introduce non-linearity, systematic heterogeneity, and mixed processes when modeling recidivism. In this research, the flexible hazard models were used to model both the criminal history accumulation process as well as the risk of post-release rearrest. Even if detailed arrest histories are unavailable to researchers, the information-theoretic approach can still be used vary profitably because it allows for very general forms for the links between attributes and the hazards.

4.1.3. Practical implications

Although much of the software needed for the analysis conducted here needed to be programmed from scratch, the availability of standard software allowing researchers to utilize information and entropy based method is increasing rapidly. For example, SAS has introduced an experimental procedure under its ETS module called PROC ENTROPY that is designed for the estimation of linear and non-linear models using the Generalized Maximum Entropy (GME) approach introduced by Golan, Judge, and Miller (1996). Additionally, LIMDEP—another popular econometrics software—has recently added the GME methods for estimating binary and multinomial logit models.

Software needed to estimate generalized hazard models using the framework described in this report here is far from being developed. In the interim, researchers and practitioners will need to rely on routines and macros developed and made available to the public. In an Appendix to this report, I have printed out the SAS macro that I wrote in order to estimate the models presented in this paper. Researchers and practitioners are welcome to copy, edit, alter, and use that code freely. However, I do not offer any performance guarantees.

Depending on the size of the sample used as well as the hardware a particular research is utilizing, the performance can vary significantly. However, the procedure is very efficient. Using the IML module of SAS (Version 8.02), a model like that of Arizona's (presented in Table 3.3) took less than 30 seconds to converge on my Dell PC (with a 3.00 GHz CPU). Note that the pre-release data in Arizona has roughly 10,000 events. Using a quarterly grid from age 0 to 100 (i.e., $\mathbf{z} = (0, 0.25, 0.50, 0.75, 1, \ldots, 99.5, 99.75, 100)'$) for the support space, this means that in each iteration, the procedure needed to evaluate a full $10,000 \times 401$ dimensional matrix. Despite that, the convergence was very rapid. With the California sample, the computer simply ran out of memory to store the matrix. Therefore, the sample needed to be truncated to 2500 individuals. This resulted in 21,838 events in the pre-release sample. For this sample, the model converged in 1.42 minutes. Therefore, despite the large sample sizes that state and local authorities may have at their disposal, the procedure should pose little or no problems on currently available computing power.

4.2. DIRECTIONS FOR FUTURE RESEARCH

The analytical framework developed in this project was not subject to simulation testing, as I concentrated on developing the framework and applying it to a substantive problem. An obvious direction for future research would be to assess the performance of the developed framework to Monte Carlo simulations. That effort would also help identify its strengths relative to existing approaches.

Another direction for future research, as was noted in the previous chapter, involves the expansion of the δ statistic to cover multiple rearrest events after release. A composite measure, aggregating the log divergence between the counterfactual and actual hazard trajectories for several rearrest events after release may (or may not) yield more clarity into the effects of incarceration.

As was noted at the end of Chapter 2, given that no model can hope to capture all unobserved heterogeneity using available attributes, it may be desirable to allow the possibility of unobserved heterogeneity via finite mixture modelling techniques.

An interesting extension of the existing approach would be to allow the simultaneous modeling of various related repeated events. For example, the framework could be extended and used to study whether and how incarceration affects the *co-evolution* of the trajectories of offending and employment (or offending and drug use) over the life course.

Finally, it would be worth utilizing the above framework for exploring how the criminal history accumulation process is deflected by other interventions (e.g., marriage, divorce, relocation, drug treatment, etc.). A large amount of society's resources are spent on trying to divert individuals from criminal offending or drug use—outcomes that have been shown to have very predictable patterns. Typically, programs designed to do so are evaluated using the standard experimental or quasi-experimental approaches. These approaches are simply different research designs used for constructing plausible counterfactuals. However, there are several instances when experimental interventions are not possible and/or when comparable control groups are impossible to find. In such settings, it may be worth investigating whether the framework developed here can be used to construct plausible counterfactuals simply by modeling the process before intervention. In a sense, one would then be evaluating the success of the intervention using *embedded counterfactuals*[1]—counterfactuals embedded in the individual's past—in order to study whether and to what extent the program was successful in achieving its goals.

[1]Historians use this terminology for a mode of reasoning that allows them to reason about the causes of historical events and actions that *cannot* be assessed using experimental or quasi-experimental approaches. See, for example, Schroeder (undated).

References

Allison, P. D. (1996) *Survival Analysis Using the SAS System: A Practical Guide*, Cary, NC: SAS Institute Inc.

Asadi, M., Ebrahimi, N., Hamedani, G. G. and Soofi, E. (2005) "Minimum Dynamic Discrimination Information Models." *Journal of Applied Probability* 42(3):643-660.

Bhorlchain, M. N. (1987) "Period Parity Progression Ratios and Birth Intervals in England and Wales, 1941-1971: A Synthetic Life Table Analysis." *Population Studies* 41(1):103-125.

Blossfeld, H., Hamerele, A. and Mayer, K. U. (1989) *Event History Analysis: Statistical Theory and Applications in the Social Sciences*, Hillsdale, NJ: Lawrence Erlbaum Association Publishers.

Blumstein, A., Cohen, J., Roth, J. A. and Visher, C. A. (1986) *Criminal Careers and "Career Criminals"*, Volumes I & II, Washington, DC: National Academy Press.

Brame, R., Bushway, S. and Paternoster, R. (2003) "Examining the prevalence of criminal desistance." *Criminology* 41:423-448.

Bureau of Justice Statistics (2002). *Recidivism of Prisoners Released in 1994, Codebook for Dataset 3355*. Downloaded from NACJD in August 2003.

Bushway, S., Brame, R. and Paternoster, R. (2004) "Connecting desistance and recidivism: Measuring changes in criminality over the lifespan." In Shadd Murana and Russ Immarigeon (eds.), *After crime and punishment: Pathways to offender reintegration*. Portland, OR: Willian Publishing.

Cressie, N. and Read, T. R. C. (1984) "Multimodel Goodness-of-Fit Tests" *Journal of the Royal Statistical Society, Ser. B* 46(3):440-464.

Donoho, D. L., Johnstone, I. M., Joch, J. C. and Stern, A. S. (1992) "Maximum Entropy and Nearly Black Objects." *Journal of the Royal Statistical Society, Ser. B* 54:41-81.

Ebrahimi, N., Habibullah, M. and Soofi, E. (1992) "Testing Exponentiality Based on Kullback-Leibler Information." *Journal of the Royal Statistical Society, Ser. B* 54(3):739-748.

Ebrahimi, N. and Soofi, E. (2003) "Static and Dynamic Information for Duration Analysis." Invited presentation given at *A Conference in Honor of Arnold Zellner: Recent*

Developments in the Theory, Method, and Application of Information and Entropy Econometrics, Washington, D.C. Accessed Feb 25, 2006 from http://www.american.edu/cas/econ/faculty/golan/golan/Papers/8_20soofi.pdf

Ebrahimi, N. and Kirmani, S. N. U. A. (1996) "A Characterization of the Proportional Hazard Model Through a Measure of Discrimination Between Two Residual Life Distributions." *Biometrika* 83(1):233-235.

Ezell, M. E., Land, K. G. and Cohen, L. E. (2002) "Modeling Multiple Failure Time Data: A Survey of Variance-Corrected Proportional Hazard Models with Empirical Applications to Arrest Data." *Sociological Methodology* 33:111-167.

Farrington, D. (1986). "Age and Crime." In *Crime and Justice* ed. N. Morris, M. Tonry 1:289-348. Chicago, IL: University of Chicago Press.

Feeney, G. and Yu, J. (1987) "Period Parity Progression Measures of Fertility in China." *Population Studies* 41(1):77-102.

Fomby, T. B. and Hill, R. C. (1997) ed. *Advances in Econometrics: Applying Maximum Entropy to Econometric Problems*, Volume 12.

Golan, A. (2002) "Information and Entropy Econometrics—Editor's View." *Journal of Econometrics* 107(1-2):1-357.

Golan, A., Judge, G. G., and Miller, D. (1996) *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, Chichester, England: John Wiley and Sons.

Good, I. J. (1963) "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables." *Annals of Mathematical Statistics* 34:911-934.

Gull, S. F. (1989) "Developments in Maximum Entropy Data Analysis." in *Maximum Entropy and Bayesian Statistics* ed. Skilling, J.Ḃoston: Kulwer.

Gull, S. F. and Danielle, G. J. (1978) "Image Reconstruction from Incomplete and Noisy Data." *Nature* 272:686-690.

Hagan, J. and Palloni, A. (1988) "Crimes and Social Events in the Life Course: Reconceiving a Criminological Controversy." *Criminology* 26(1):87-100.

Harding, R. W. and Maller, R. A. (1997) "An Improved Methodology for Analyzing Age-Arrest Profiles: Application to a Western Australian Offender Population." *Journal of Quantitative Criminology* 13(4):349-372.

Hendry, D. F. (1995) *Dynamic Econometrics*, New Your, NY: Oxford University Press.

Hinde, A. (1998) *Demographic Methods*. New York, NY: Oxford University Press Inc.

Hirschi, T. and Gottfredson, M. R. (1983) "Age and the Explanation of Crime." *American Journal of Sociology* 89:552-584.

Huber, P. J. (1967) "The Behavior of Maximum Likelihood Estimators under Non-Standard Conditions." *Proceedings of the Fifth Symposium on Mathematical Statistics and Probability* 1:221-233.

Jaynes, E. T. (1957a) "Information Theory and Statistical Mechanics." *Physics Review* 106:620–630.

Jaynes, E. T. (1957b) "Information Theory and Statistical Mechanics II." *Physics Review* 108:171–190.

Jaynes, E. T. (1982) "On The Rationale of Maximum Entropy Methods." *Proceedings of the IEEE* 70(9):939-952.

Journal of Econometrics: Special Issues on Information and Entropy Econometrics (2002). ed. Amos Golan. 107(1-2): 1–357.

Kullback, J. (1959) *Information Theory and Statistics*. New York, NY: John Wiley.

Land, K. and Nagin, D. (1996) "Micro-Models of Criminal Careers: A Synthesis of the Criminal Career and Life Course Approaches via Semiparametric Mixed Poisson Regression Models." *Journal of Quantitative Criminology* 12(2):163-191.

Langan, P. A. and Levin, D. J. (2002) *Recidivism of Prisoners Released in 1994*. Special Report, Washington, DC: Bureau of Justice Statistics.

Levine R. D. (1980) "An Information Theoretic Approach to Inversion Problems." *Journal of Physics A* 13:91-108.

Maasoumi, E. (1993) "A Compendium of Information Theory in Economics and Econometrics." *Econometric Reviews* 12(2):137–181.

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

Maltz, M. and Mullany, J. M. (2000) "Visualizing Lives: New Pathways for Analyzing Life Course Trajectories." *Journal of Quantitative Criminology* 16(2):255-281.

Mayer, K. U. and Tuma, N. B. (ed.) (1990) *Event History Analysis in Life Course Research*, Madison, WI: The University of Wisconsin Press.

Mittelhammer, R. C., Judge, G. G., and Miller, D. J. (2000). *Econometric Foundations*, Cambridge, UK: Cambridge University Press.

Nagin, D. S. (2005) *Group Based Models of Development*, Boston, MA: Harvard University Press.

Nagin, D. S. and Land, K. (1993) "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation on a Non-Parametric, Mixed Poisson Model." *Criminology* 31:327-362.

Piquero, A. R., Farrington, D. P., and Blumstein, Al. (2003) "The Criminal Career Paradigm." In *Crime and Justice*. ed. M. Tonry 30:359-506. Chicago, IL: University of Chicago Press.

Ryu, H. K. (1993) "Maximum Entropy Estimation of Density and Regression Functions." *Journal of Econometrics* 56:397-440.

Sampson, R. J. and Laub, J. H. (1990) "Crime and Deviance Over the Life Course: The Salience of Adult Social Bonds." *American Sociological Review* 55:609-627.

Sampson, R. J. and Laub, J. H. (1993) *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.

Schmidt, P. and Witte, A. D. (1988) *Predicting Recidivism Using Survival Models*. New York: Springer-Verlag.

Schroeder, P. W. (undated) "Embedded Counterfactuals and World War I as an Unavoidable War." forthcoming in *Unmaking the West: Counterfactual Thought Experiments in History*, edited by Philip Tetlock, Richard Ned Lebow and Geoffrey Parker. Accessed Feb 25, 2006, from http://www.asu.edu/clas/polisci/cqrm/papers/schroedercounterfactual.pdf

Shannon, C. E. (1948) "A Mathematical Theory of Communication." *Bell System Technical Journal* 27:379-423.

Soofi, E. S. (1994) "Capturing the Intangible Concept of Information" *Journal of the American Statistical Association* 89(428):1243-1254.

Soofi, E. S., Ebrahimi, N. and Habibullah, M. (1995) "Information Distinguishability with Application to Analysis of Failure Data." *Journal of the American Statistical Association* 90(430):657-668.

Visher, C. A., Lattimore, P. K., and Linster, R. L. (1991) "Predicting the Recidivism of Serious Youthful Offenders using Survival Models." *Criminology* 29(3):329-362.

White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48:817-830.

Zellner, A. (1991) "Bayesian Methods and Entropy in Economics and Econometrics." In *Maximum Entropy and Bayesian Methods*, ed. W. T. Grady and L. H. Schick, 17-31, Netherlands, Kulwer.

Zellner, A., and Highfield, R. A. (1988) "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions." *Journal of Econometrics* 37:195-209.