DIFFERENTLY RATIONAL
ESSAYS ON CRIMINAL BEHAVIOR

BY GEOFFREY FAIN WILLIAMS

A dissertation submitted to the

Graduate School - New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Tomas Sjöström

and approved by

____

____

____

____

New Brunswick, New Jersey

October, 2012

ABSTRACT OF THE DISSERTATION

Differently Rational
Essays on Criminal Behavior

by Geoffrey Fain Williams

Dissertation Director: Tomas Sjöström

This dissertation uses a range of economic tools to analyze and understand criminal behavior, particularly theft. The first chapter outlines a number of key issues and stylized facts observed in criminal behavior, and provides an outline of the chapters that follow.

The second chapter proposes a simple threshold model of theft, and develops a number of structural estimators based on this model. It then tests the model against data from the National Longitudinal Study of Youth, 1997 Cohort. The evidence suggests that the key determinant of theft behavior is the costs of theft to the thief, and in particular the thief's perception of future costs. There does not seem to be significant variation in the benefits of theft; that is, there is no sign that some individuals are more capable of theft than others. The data also shows that theft behavior is usually very short-lived, with the vast majority of thieves showing activity for less than one year in adolescence.

The third chapter looks at the temporal pattern of criminal behavior (frequently termed a "criminal career" or "trajectory") in individuals' lives, up to the age of 25-30. It uses three different data sets, based on several methods of observation, and finds a number of similarities. In contrast to earlier work describing criminal careers, the data suggest that the two measures of age-specific inclination and individual-specific intensity are the key to describing patterns in criminal behavior. Specifically, there is significant evidence that an individual A committing a single crime at age 14 has more in common with individual B who commits only one crime, at age 24, than C who commits three crimes at age 14.

The fourth chapter looks at the decision to steal in the context of a simple model of human capital accumulation, as a way to tease out the relative role of labor substitution and impatience in individuals' decision to steal. The data support the role of impatience a significant driver. First, individuals who report stealing show a wage and labor market participation pattern strongly consistent with a low discount rate. Second, individuals who report stealing show significant underinvestment in education, with lower enrollment rates and higher grade repetition than comparable non-thieves. Finally, individuals who report stealing show a larger number of employers, suggesting underinvestment in long-term career success.

Acknowledgements

By the end of my first semester in the program, I had developed a solid understanding of my situation. "It's like jumping off a building", I told friends, "You don't really need to worry about the first part - with the right attitude, it's guaranteed to be a lot of fun. The *really* crucial issue is how you land."

So after five years, having landed in a tenure-track position (with no bones broken), and having had fun the whole time, I feel very fortunate indeed. The time has come to thank the small army of people who provided guidance, support and advice throughout and helped contribute to this good fortune and good fun.

The list of course begins with my committee, who patiently watched and coaxed me as I tried to take an inchoate hunch about crime data and turn it into real research. I am deeply indebted to them, and very glad to know them.

In the fall of 2008 I took the course that Rutgers Economics calls IO II with Professor Tomas Sjöström . I didn't learn anything about industrial organization (and in fact I don't think that topic ever actually came up in the course). Instead, Professor Sjöström led us through a fascinating extended meditation on the limits of human cooperation, in a course that would have been better titled "Mechanism Design and the Fall of Man." It was one of the most enjoyable courses I've ever taken, and changed my view of the world. His wide-ranging interests, his openness to helping develop any idea, and his creative insights all made an enormous difference over the past few years.

Professor Anne Morrison Piehl was the first person to express an unqualified enthusiasm for my potential as a researcher, something for which I will always be grateful. She provided me with an amazing introduction to the issues and personalities involved in the analysis of crime. Her remarkable intuition and insight and her lightning-fast grasp of the key issues were crucial resources at every stage. But her readiness to cheerfully call me out on my overconfidence or other foibles is what I found most enjoyable, and what I am most grateful for.

Professor Roger Klein was remarkably patient as I got my mind around some of the key issues of econometrics. His sincere concern for me and all his students and his dedication to his work and the field as a whole are an inspiration and a model to me. One of my proudest accomplishments of the past few years is moving from the confused student to a very junior colleague who is able to contribute the occasional real insight when we discuss econometric issues.

Several of the key papers that helped me to structure my thinking about crime were authored or co-authored by Professor Lance Lochner, presently at the University of Western Ontario. I came to realize he was uniquely well-suited to evaluate my work. I was very grateful that he was willing to be the outside reader, and even more grateful for the extensive thoughtful comments he gave each of the chapters, which greatly focused and deepened my approach.

Although this is my formal committee, I feel as if I have a “shadow committee” that is many many times larger. The Rutgers economics faculty as a whole has been extremely supportive. When I added a list of faculty members who had helped me to one paper, a classmate who read it exclaimed “You’ve been bothering so many people!”. And the sad thing is it was really only a partial list. Virtually every single professor gave me a hand at one point or another. However, Richard McLean and Marty Perry deserve special attention for the countless ways they helped me and all my classmates along.

Words fail me when describing my debt to Dorothy Rinaldi. I find it difficult to believe

New Jersey Hall is still standing after her retirement. The good-humored assistance of Janet Budge, Donna Ghilino, Janet Goodstein, Paula Seltzer, Debra Holman and particularly Mayra Lopez were vital at all stages of my time at Rutgers.

My classmates have also been uniformly friendly and supportive, but Levent Ulku, Junnosuke Shino and Costanza Biavaschi stand out as integral to my success (and assistance by Georgia Bush, Ebru Isgin and Steve Gazzillo was greatly appreciated).

Before I got to Rutgers, my way was smoothed by a number of people - Yaw Nyarko, William Easterly, Felix Oberholzer-Gee, Charles Wilson, Paul Kleindorfer, Ashley Timmer and Anjani Jain gave me critical assistance as I approached the idea of a doctoral program.

A very incomplete list of friends who assisted would still have to include David Rothschild, Mike Damron, and Thomas Lurquin, who listened to my hopes and fears, offered advice and gave me pep talks at uncountably many key moments over the past decade. Daniel Martin's gravity-defying upbeat attitude and incredibly generous nature has to be singled out for praise - I am writing this on a desktop computer he gave me.

The idiosyncratic utility function that made an Economics doctorate a rational decision I owe to my family, particularly my parents. I remember when I was 17 my mother told me "I think you'd enjoy being an economist - you like numbers and you like people." Being the attentive and dutiful son that I am, it was only two decades later that I began applying to programs (whatever else a reader may take away from this dissertation in terms of structural models or criminal career paths, I hope you gain a renewed appreciation of the vital importance of Listening To Your Mother). My father's passion for understanding the theoretical underpinnings of social organization spilled over from an early age. The passion that both of them have for knowledge and fairness have always been inspirational to me, and I cannot think of any aspect of this dissertation or of my life that was not strongly influenced by them.

My siblings and extended family have also been wonderfully supportive and inspirational, and I am grateful.

Nothing in this document would have been possible without the calm presence and childcare assistance of my mother-in-law, Sandra Raupfer.

And finally, I would like to thank my wife, Chrysantha Clark. For the past 5 years I have not just been Chrysantha Clark's fellow, but also a Chrysantha Clark Fellow. Her support, engagement, assistance and tolerance made a difference every day that I worked on this. It is to her and our son Matthias that I dedicate my dissertation.

Dedication

To my wife Chrysantha Sophia Clark, and our son, Matthias, who were so nice to come home to.

He has sounded forth the trumpet that shall never call retreat;
He is sifting out the hearts of men before His judgment-seat:
Oh, be swift, my soul, to answer Him! be jubilant, my feet!
Our God is marching on.

Fourth stanza, The Battle Hymn of the Republic, Julia Ward Howe.

Table of Contents

List of Tables

List of Figures

Chapter 1

Three Players, Infinite Periods

1.1 Introduction

If we can understand why some men steal, this knowledge can potentially yield immediate dividends by helping us to control and prevent crime. But there is an even greater potential benefit over the long term, for once we understand why some men steal we will have a much clearer understanding of why other men *do not* steal, and thence of the fundamental bindings that hold together society.

The following essays look at several aspects of criminal behavior and try to understand how the patterns of criminal behavior can be incorporated into a larger model of humans as optimizing agents, using the toolkit of economics.

The first essay "A Simple Threshold Model of Theft" proposes and estimates a simple threshold model of theft, where all individuals have access to the same distribution of theft opportunities, and what differentiates them is their expectation of the cost of theft. The data support the basic assumptions of the model. The best predictors of theft are variables that are good proxies for patience and forward thinking.

The second essay "Criminal Careers In Two Dimensions" looks at data on criminal behavior from several well-established data sets and finds a high level of randomness at the individual level, that nonetheless follows fairly simple patterns when aggregated. It reinforces the general finding in criminology of the importance of the age-curve - the steady increase in criminal activity to a peak at age 17, with a slow fall from that point.

The third essay "The Thief's Wages" develops a variant of the model in the first essay, and integrates the decision to steal into a larger model of human capital development. As with the first essay, it finds that the best fit of the data with the model is provided by allowing thieves to be systematically more impatient than non-thieves.

A few general patterns come through in all three essays that are consistent with patterns found by previous researchers. First, covariates we would associate with impatience and impulsivity, specifically smoking behavior, emerge as strong predictors of crime. Secondly, crime is clearly focused around adolescence. Third, direct measures of punishments, costs and benefits (i.e., wages) do not appear to be very powerful in understanding the decision to steal.

There are two general points that I believe help us to understand crime in this way, and see how these and other patterns in the data can be fit together.

First, the most important costs that a criminal faces for committing crime are costs inflicted informally by the people around him. That is, the most intense costs of crime to the criminal are the costs inflicted by society, not the state.

Second, every individual faces a wide range of opportunities to commit crime, with widely variable benefits. For each individual, most of these opportunities are very low benefit - stealing a pack of gum, say - but a few are extremely lucrative - embezzling from the company's accounts, for example. The variation *within* the pool of benefits available to each individual is far greater than the variation in the explicit costs of "being a criminal" *between* individuals (that is, the variation in the probability of being caught, the opportunity cost of time, the future impact on earnings of gaining a criminal record, etc.). Thus, while it seems probable that directly measurable costs and variation in marginal utility of money play some role in the decision to commit a particular crime, it is highly unlikely that they play a major role in the decision to commit crime in general.

In the remainder of this introduction I will attempt to establish both assertions.

1.2 Society acts last and most powerfully

Two examples can help to illustrate the power of informal social action.

Abani (2004) tells the story of an incident in a local market in Nigeria, most likely in Afikpo, in the 1970s, where an unnamed man was accused of stealing from a vendor in a local market. The theft was not proven, and the details of what was stolen are unclear, but most likely it was food or a few items of fairly low value, perhaps a few days' wages.

Nigeria, like many other African countries, has a system where local communities are generally governed by customary law, with decisions made and enforced by local representatives. This exists alongside a more centralized system of common law; cases can be moved from the customary law system to the common law system. Afikpo is a traditionally Igbo area, and the report on the crime and its aftermath comes from an individual with an Igbo father; while I do not have direct proof, most likely Igbo customary law would have had precedence in any criminal proceeding. While penalties for theft under Igbo customary law vary across different locations and time periods, punishments of another Igbo community in the 1980s and 90s were limited to fines and public humiliation (for a general discussion of customary Igbo law and its place in a community see Okereafoezeke (2003)).

Wedgwood (1962) briefly relates a story from the England of Charles I. In the 1630's, as the country moved towards the religio-political crisis of the Civil War, three Puritans published pamphlets and made public statements criticizing a group of bishops close to the king.

Because many of the attacks came close to the king himself, this was considered the offense of sedition, and punishments included fines, prison and mutilation. In the mid-1500s it might have been construed as treason, which could be punished with death (for a discussion of the idea of sedition and its development from the 1500s to the 1650s, see Manning (1980)).

In the case of the Nigerian thief, the potential payoff relative to the expected cost of the formal punishment is fairly high, especially if the thief thought he would be likely to avoid detection. The theft seems to be a relatively good bet; looking purely at the material benefits and costs of the immediate situation and the legal context, and ignoring any ethical issues, we can see why this might be a reasonable action.

In the case of the English Puritans, there is no clear material payoff to denouncing the King's policies, the crime was highly public and the odds of strict formal punishment high. Looking purely at the material costs and benefits and the legal context, it's unclear why someone might take such a risk.

In the incident in 1970s Nigeria, the theft was discovered by some of the market vendors. The outcome was vividly and horrifyingly described by the Nigerian writer Chris Abani in the New York Times in 2004[1]:

> Suddenly, a lone voice screamed one word over and over: "Thief! Thief! Thief!" It was picked up slowly, as if the drizzle that afternoon had dampened the scent of blood. My aunt froze and faced the sound, nose sniffing the air like a lioness sensing prey. A man's voice, tired and breathless, tried to counter the rising chant with a feeble retort: "It's a lie! It's a lie!"
>
> ... The sound of a chase grew closer – desperate, pounding feet and shouts. A man with a wild expression came around a corner and nearly knocked over a stall. Young men were almost on his heels, followed some distance behind by a larger cluster. The man ran past us, and it seemed as if I could see the pores in his skin. ...[H]e was headed for the courtyard in the middle of the market, where the retired elders sat daily to dispense justice...The council of elders was the highest court in my community, besides, of course, the civil system. It arbitrated on everything from murder to marital disputes, and its authority was never questioned.
>
> ... The man stood in the middle of the clearing facing the elders while the crush of people pressed around them. In the center of this sacred space, the sole elder to

[1] I came across a quotation from the account in Thurston (2011), which includes other examples of similar mob violence in a range of cultures and periods.

> stand up and call for tolerance was booed and pelted with rotten fruit. He sat down quickly and turned his face away. I was sure that the man was about to be lynched. How could the crowd ignore the elder's intervention? And why didn't the other elders speak out?
>
> The mob was oddly silent; its loud breathing filled the space. The accused man began to beg, but people were too busy picking up stones and tree branches, anything that could be used as a weapon. A young man broke through the crowd carrying an old rubber tire and a metal can. He hung the tire from the accused's neck. This singular action ended the man's pleas for mercy. Resigned, he sobbed softly, mumbling inaudibly, but he didn't move as the young man emptied the contents of the can onto him. The young man smiled and talked as he went about his task: "You see why crime doesn't pay? I am doing this for you, you know. If you burn here, you won't burn in hell. God is reasonable."
>
> Finishing, he held up a box of matches. The crowd roared. The elder who had tried intervening spoke again, but nobody listened. Someone called out, "Bring the children forward so that they can learn." My aunt hustled me to the front.... I never saw the match fall, but I felt the heat as the man erupted into a sheet of flame, burning like a lighthouse in the drizzly haze. "Watch," my aunt said as I tried to turn away from the writhing figure.

In the England of 1637, the state acted directly to judge and punish the three Puritans Prynne, Bastwick and Burton. As summarized by Wedgwood (1962): "[A]ll three were duly brought before the Star Chamber and sentenced. All three were pilloried, lost their ears[2], suffered huge fines, and were sent to prison for life."

As harsh as this is, the description of the execution and its aftermath suggest that their actions may have been more rational than a simple formal description acknowledges:

> The people...greeted them in the pillory with acclamation; Bastwick's wife carefully received her husband's ears in a clean handkerchief, laughing and joking with him the while, and when the executioner had done with him, climbed upon a stool, put her arms about his neck as he stood in the pillory and kissed him on both cheeks, the people loudly applauding. When the prisoners left London the streets were lined with sympathetic onlookers who showered gifts and good wishes upon them, and at Chester the Mayor and Corporations welcomed Prynne with a civic dinner and a gift of hangings for lodgings in prison.

It is important to note that both situations, while extreme, are far from unique. Instances of mob violence similar to the situation in Nigeria have been observed in virtually

[2] Most of Prynne's ears had been cut off several years earlier, for earlier pamphlets, so the executioner removed the remainder (Cheyney, 1913).

every society across every region and every century. The rewards provided to the three Puritans by their countrymen can also be observed in a number of instances. I chose the two stories for two reasons: first, the vividness of the descriptions, and second, the fact in both cases the available information strongly shows the punishments were inflicted purely for actions, not identities (mob violence of course frequently has an ethnic, religious or racial component).

1.2.1 Three players, infinite periods

In analyzing an individual's decision to commit or not commit a crime, we can usefully simplify a range of complicated situations by focusing on three "agents" - the individual who has the choice to offend, the formal state structure that has jurisdiction in the investigation and prosecution of the crime, and the informal social structures that surround the individual. I will abbreviate these three as the individual, the state and society (victims and co-offenders may change the details, but will not affect the overall issues).

The stories above reinforce a general point that is sometimes not given the attention it deserves in the analysis of criminal behavior: society tends to get the last move, and tends to have the most power.

In both cases described above the formal systems of state power were secondary to the response of the informal systems that surrounded the individual, and in the case of the Nigerian thief, the formal systems, such as the community elders, simply sat by the sidelines.

The fact that informal social sanctions play a huge role in the individual's expected utility from either committing or desisting from a crime would suggest that greater understanding of these sanctions is important. But it is quite difficult to find much empirical work that directly focuses on these informal sanctions. In economics there is a substantial theoretical literature on cooperation among agents (either individuals or firms), most of the empirical work is focused at issues of collusion among firms. There is a related literature in biology on cooperation. Both literatures have experimental work behind them, but little analysis of real-world situations. In anthropology many (perhaps most) ethnographies have at least a snippet or two that describes what happens to individuals who break the mores of the society being described. I have only been able to find one detailed systematic effort to put together a description of the informal penalties that people incur for breaking the rules of their social world: Merry (1984).

As Merry (1984) makes clear, social groups can enforce their rules using wide complement of tools, ranging from the sweetest carrots to the most violent of sticks. She focuses part of her discussion on the Sarakatsan shepherds in Greece. Among the Sarakatsan, there is a fairly public and uniform assessment of every family and every individual that is summarized in their "honor." To have a high level of honor has consistent and ongoing material rewards such as access to pasture lands. Any loss of honor threatens these rewards and can also lose other privileges. Across the range of ethnographies she reviews, to lose favor among the other members of your social group is extremely risky. An individual who violates social rules regularly and/or egregiously goes from a loss of material rewards to greater and greater humiliations and stigma. Across the range of societies described, individuals who fail to follow the rules of the social group are at risk of actual violence, ostracism and even death[3].

A full specification of the repeated interaction of these three agents - the individual, the state, and society - is beyond the scope of this paper, but using established results of game theory we can sketch out some of the issues. I am thinking in particular of the literature on infinitely repeated games.

It seems reasonable as a basic model that the individual has relatively little power to affect the payoff of either society or the state, while conversely both society and the state

[3] I am skipping over many of the details and subtleties of this excellent article. Readers with an interest are urged to read the original.

can drastically affect the utility of the individual with relative ease. If the individual's behavior is "good" (honest, cooperative, law-abiding) that yields benefits to society and the state, but if it is "bad" (dishonest, violent, uncooperative) than a range of penalties can be enforced at low cost. One way this might be described theoretically is a stage game where the difference between the individual's minmax payoff and his other payoffs was very substantial, while none of the payoffs of society or the state varied significantly.

Additionally, based on the discussion in Merry (1984), we can imagine society being able to subtly vary the per-period payoffs of the individual, based on his compliance levels. The payoff in a single time period for "excellent" behavior could be just a bit higher than the reward for "very good" behavior, for example. Over multiple periods, these subtle differences will aggregate into very substantial differences, giving society remarkable power.

Adding to society's power in this repeated game is the potential to inflict a grim trigger strategy on the individual. And the grim trigger need not be anywhere near as grim as the lynching Abani describes. By ostracizing the individual for a serious infraction (a penalty that many societies enforce against serious offenders), society can effectively reduce the individual's payoff to zero for the remainder of play. When compared to the aggregated benefits of positive rewards across an indefinite number of periods, a zero payoff is an extremely harsh reduction in utility.

In comparison with society, the state has some of the same powers, but is generally more limited. It has less ability to monitor the individual than his friends, co-workers and neighbors. It is likely to have fewer carrots to offer the individual for good behavior, and is likely to have more trouble enforcing any punishment or reward over long periods of time. Most of its rewards and punishments will be fairly uncertain and blunt, such as temporarily detaining an individual.

1.3 Variation in opportunities vs variation in costs

1.3.1 Variation in the benefits of crime

One of the stronger patterns that has come from research into the earnings from crime has been the wide range of variation.

In a study of active drug dealers in Washington, D.C., Reuter et al. (1990) find that "reported earnings were very skewed, reflecting the great range of selling frequency." Dealers who sold on a daily basis earned a median of \$2,000 per month net of any costs (the mean net income for daily sellers was \$3,600, reflecting substantial variation even within this group). Dealers who sold only one day a week or less earned a median monthly net income of \$50 (mean \$160, again reflecting significant variation).

While Wilson and Abrahamse (1992) focus on the interquartile range in their data on the self-reported crimes and earnings of prison inmates, they find enormous range in number of acts of crime which immediately leads to dramatic range in earnings from crime. They write, "it makes little sense to analyze the costs and benefits of crime from the view of the average offender; for all practical purposes there is no such person."

Re-examining the same data, Tremblay and Morselli (2000) include a wider range of reported activity and earnings levels. They find that about 15% of inmates average fairly high annual earnings of \$4,566, while another 31% of inmates are both less active and less efficient and earn only \$211 per year from crime on average.

In the first chapter of this dissertation, I look at earnings from theft for individuals in the NLSY and find a similar wide variation on both an annual basis and on a per-act basis. The median per-act earnings is roughly \$30, and the range goes from zero dollars to over five thousand dollars.

In summary, across a range of data sets, the earnings from crime show enormous range, with the high earners earning a minimum of 40 times the low earners. There is no evidence in any of these data sets that the high earners are particularly capable relative to the general population; those earning $4,566 per year from crime are not supercriminals. All of the researchers seem to suggest that a wide range of noncriminals could achieve the same earnings (as Wilson and Abrahamse (1992) write, "[t]he wonder is that more people don't steal")

1.3.2 Variation in the costs or marginal utility of crime

When the range of realized opportunities vary by a factor of 40, and there is no evidence that variation in ability explains much of this, it would seem logical that variation in costs, or in marginal utility, explain why some people become criminals for $200 and others hold out for $4,000.

A very common explanation for the fact that some people steal and others don't is income or wealth; the thieves are poorer and need the money more (an alternative, economist variant of this, is that thieves have lower opportunity costs, see Grogger (1998) or Freeman (1999)). As a conjecture, this is quite logical, and a number of interesting papers have found some evidence of it (Freeman (1999), Freeman (1991), Grogger (1998) Lochner (2004), and Gould et al. (2002)). However, there is a distinct lack of any smoking gun.

When we compare data on earnings from crime with data on variation in consumption, the idea starts to look even weaker. The recent paper showing the strongest consumption inequality in U.S. data is Aguiar and Bils (2011). Comparing consumption levels in the 5th to 20th percentile of income with consumption levels in the 80th to 95th percentile from 1980 to 2007, they find that the ratio of the consumption of the two groups has fluctuated between 2.5 to 3. That is, the variation in earnings among thieves is something like 10-15 times the variation in consumption between the wealthy and the poor in the United States. It would make sense that an individual at the 80th percentile would hold out for an opportunity 3 times greater than an individual at the 20th percentile, but if the criminals from poorer households were only earnings \$200 in late 80's, this would mean the wealthy should hold out for any opportunity over \$600. Put differently, if \$4,000 was far too low to entice someone at the 80th percentile, than essentially nobody should have been breaking the law for anything less than \$1,000 per year.

1.4 Putting these together

I have argued that informal social groups that surround the individual have the power to offer rewards for compliance and punishments for offenses, and that a vital part of that power comes from subtle gradations in payoffs over along periods, a number of stylized facts of criminal behavior become clearer. Additionally, I point out that variation in potential opportunities from crime are far greater than variation in the directly measurable costs of criminal behavior to the criminal.

Putting these two ideas together, a number of stylized facts of crime become easier to understand.

The vast majority of crime is committed by individuals between ages 15 and 25. To the degree that an individual has the option to move to another location or social setting, it drastically reduces the power of society to affect his payoffs. The freedom to move away will be at its highest in adolescence, when the individual is no longer dependent on his family of origin, but does not have an established role in the larger group. Misbehavior will thus be "cheapest" at this point, and thus most common.

Researchers consistently point to self-control and impulsivity issues as causes of crime.[4]. Since for an individual choosing whether or not to commit a crime all punishments are in the future, the more awareness he has of future consequences, and the less he discounts future outcomes, the costs of punishment will be lower for more impatient individuals. The fact that the most powerful rewards and punishments will be implemented

[4] I review the evidence for this in sections 2.1 and 4.2.

by society over long periods of time means that minor differences in patience will lead to radical differences in how the individual experiences the cost of punishment.

Evidence suggests that the majority of youthful offenders (i.e., the majority of offenders) commit crime only or largely with accomplices.[5] While I have discussed society as a monolithic entity to simplify matters, it is of course composed of individuals. The judgment of peers will be part of the overall judgment of society. For an adolescent, his peers will be the group he can expect the most extended contact with - older individuals will die off earlier, younger individuals are not completely developed and so less "knowable". Thus, if other members of his peer group are taking part, that substantially lower the costs of committing a crime.

It is very difficult to find strong effects of measurable, formal costs and benefits of criminal behavior. A standard model for analyzing the decision to commit a crime looks at only at the formal expected cost of capture and punishment. Thus, to steal might be modeled as incurring an expected cost of $E(P) = p \times c$ where p is the odds of capture and c is the disutility of the punishment inflicted by the state. In a wide range of cases p is likely to be substantially below 1/2, and if c is a brief stint in prison or less, then $E(P)$ can be quite low. As Wilson and Abrahamse (1992) comment, "[t]he wonder is that more people don't steal." But if we factor in the punishments that society can resort to (such as that experienced by the Nigerian thief), $E(P)$ becomes a much more complicated but much more powerful expression.

1.5 Does this mean that material costs and benefits have no explanatory power?

To say that direct material benefits and costs have limited explanatory power is not to say they have no explanatory power. There are a number of ways that direct material benefits and costs can enter into the decision to commit crime.

[5] This is reviewed in section 2.1.

First, the benefits and costs of committing a crime most likely affect the decision to commit a crime at the margin. In my first chapter, I find evidence that more patient thieves hold out for higher earnings from crime; implicitly, the model suggests that lower expected returns would not have enticed the same thieves. However, this variation in "price" seems likely to be a small part of the overall equation.

An analogy can be made with the issue of "copycat" behavior. A significant literature has shown that in the wake of coverage of suicides, suicide rates increase (Phillips and Carstensen, 1986). That is, the public discussion of a single suicide increases the odds of another individual committing suicide. It seems possible that the public example lowers the social cost to somebody who is already thinking of suicide. However, the overall effect, while significant, is small; the change in suicide rates can only be established with large datasets. Similarly, the change in the "price" of the crime or the cost of formal punishment enter into the decision, but as small parts of the overall equation.

Secondly, a situation where material benefits of crime are likely to be substantial would be instances of systematic, ongoing corruption. In these cases, some of the social groups surrounding the individual are complicit in the crime. The rewards of corruption become part of the repeated payoffs these social groups can offer the individual (and at the same time, these groups can inflict intense punishments). But in contrast to crimes that are committed by small isolated groups, corruption is more longlived, and the payoffs can be sustained for longer periods.

In a more general sense, there is some reason to think that measures of income and wealth could be negatively correlated with criminal activity.

First, the work of Heckman and his students (for example Cunha et al. (2006) and Lochner and Moretti (2004) has provided persuasive evidence that more education and earlier education can help to increase patience levels. More income and wealth yields greater resources for education and early childhood development, yields lower levels of criminal activity.

Second, while I have not been able to investigate the issue of social structures directly in the attached essays, there is substantial evidence that neighborhood structure and cohesiveness affect criminal behavior (Sampson et al., 1997). While collective efficacy is only partially determined by income and wealth, these variables play a role.

1.6 Conclusion

The preceding discussion has tried to sketch out some general ideas. The remainder of this dissertation looks at three specific issues and tries to support several novel interpretations of the data on crime.

The first essay gives a simple model of theft. It argues that there is little difference in ability across thieves, and between thieves and non-thieves. What explains differences in theft behavior is not differences in benefits, but differences in the costs of theft. The evidence suggests that perceived differences in cost are particularly important.

The second essay looks at the patterns of criminal behavior from the early teen years to the mid to late twenties, using a number of data sets. It points to the highly stochastic nature of that behavior, but also several consistencies that appear to apply across a range of data sets. The level of a criminal's intensity appears to explain the number of acts of crime he commits, but there are lengthy periods where all but the most active criminals are not active.

The third essay applies a variant of the model used in the first essay, and links it to a larger model of human capital acquisition. The data add further support to the idea that impatience appears to play a major role in the decision to commit crime.

Chapter 2

A Simple Threshold Model of Theft

2.1 Introduction

The use of economic tools to investigate crime has been a significant field of research since Becker (1968)[1]. Despite the success of the approach there are a number of factors that continue to show a deeper and more complicated role in criminal behavior than would be easily predicted by a rational agent model. Particularly intriguing are the empirical relationships between crime and adolescence, crime and and patience and self-control, and crime and peer effects.

Criminal activity is particularly intense during adolescence, such that delinquency has been described as apparently a "normal part of teen life"[2]. The pattern is almost perfectly consistent across regions and time periods, such that the relative change in activity across ages (absolute rates in the population vary) looks effectively identical in data from such disparate places as the contemporary United States, Argentina in the 1960s, and in the England and Wales of the early 1840s[3].

Criminals appear to consistently downweight future consequences. There is a significant

---

[1] See Freeman (1999) for a summary of work up to 1999

[2] The quote is from Moffitt (1993) p. 675 which disaggregates delinquents (with follow-up work by Nagin and others) into the two groups of *life-course-persistent* criminals and *adolescent-limited* criminals.

[3] From Hirschi and Gottfredson (1983), who remark "If the form of the age distribution differs from time to time and from place to place, we have been unable to find evidence of this fact." Economists have found some evidence that lower opportunity cost and punishment risk of juveniles can help to explain the difference, but the evidence is mixed (see Levitt and Lochner (2001), Levitt (1997), Lee and McCrary (2005), Grogger (1998)) and there is a generalized pattern of greater risk-taking in adolescence that does not seem to have a direct economic explanation (see Spear (2000), Guo et al. (2010)).

body of literature in criminology arguing that impatience and poor self-control are absolutely critical to understanding crime (Gottfredson and Hirschi, 1990). Interventions that increase lifelong patience and self-control have shown decreases in both criminal behavior and other risky choices. Participants who were randomly assigned to the High/Scope Perry Preschool Program were arrested at 28% of the rate of the control group, and were 23% less like to smoke than the control group. This was despite the failure of the program to effect a long-term change in academic achievement tests[4].

While difficult to define and measure (Manski, 1993), there is ample anecdotal and quantitative evidence of peer effects in crime. In a data set on youth offenders more than 87% of all robbers and more than half of all offenders acted with at least one confederate[5].

In this paper I propose a simple threshold model of theft, and test it against data from the NLSY 1997 Cohort and empirical findings by other researchers. In the model, the value of theft opportunities are lognormally distributed, where the distribution has the same parameters for all individuals, and on a daily basis each individual has access to one opportunity. Each individual has a cost "threshold", which represents the expected cost of stealing to them. If the opportunity presents a greater expected benefit to them than the cost, they take it, otherwise not[6].

---

[4]The Perry data is from Belfield et al. (2006). Heckman and his collaborators have found a strong negative relationship between noncognitive skills (feelings of control and self-worth) and criminal behavior as well as other risky behaviors such as smoking (Heckman et al. (2006), Cunha et al. (2010)). While not a favored explanation, economists have referenced both impatience and lack of self control as possible explanatory factors in the analysis of crime (Lee and McCrary (2005), Levitt and Venkatesh (2000)).

[5]Data from Zimring (1998). For a range of discussions of peer effects see Case and Katz (1991), Glaeser et al. (1996), Kling et al. (2005). It should be noted that economic analyses have tended to define peer effects using what might be called a "contagion" model, where a more criminally-inclined individual influences a less inclined individual. In contrast, other disciplines have discussed what might be called a "reverberation" model, where equally risk-averse individuals induce greater risk taking in each other (Gardner and Steinberg, 2005).

[6]There are a number of models of crime from other disciplines, as well as economic models focused on other problems, that resemble this theory. Within economics it bears some resemblance to the search models of labor economics. It also has some similarities to the foraging models from ecology that have helped to inform neuroeconomics (Glimcher, 2004). Three papers from criminology and the economics of crime that touch on some ideas are Lee and McCrary (2005), Grogger (1998) and Morselli and Tremblay (2004). Marcus Felson's theory of routine activities is similar in its focus on the costs and benefits of specific criminal opportunities (Felson, 2009).

Examining the annual data from the NLSY reveals a pattern that is consistent with other research but does not appear to have been noted: theft behavior is extremely spiky, with a median theft career length of less than a year. Given the limitations of the data it is impossible to estimate the length of the average career, but six to nine months seems roughly correct. The model accommodates this pattern with the assumption that individual cost thresholds shift at intervals.

I test the model against the data in a number of ways, including the use of simulations and a range of parametric and semiparametric estimations. I demonstrate that there are a number of general patterns in the data that are most parsimoniously explained using the model. Additionally I estimate two structural models, which I evaluate in two ways.

First, I compare the distribution of marginal effects across a range of specifications, including parametric and semiparametric specifications of both binary discrete choice and count data. Structural specifications based on the model show marginal effects consistent with each other and with the semiparametric estimates, while several standard parametric specifications[7] completely fail on this test.

Second, I look at the significance and explanatory power (measured via pseudo-R^2) of a large number of covariates. While it is not clear a priori what covariates would predict greater opportunities or ability to commit theft, regional and locational dummies show no power. Measures of economic cost (for example, opportunity cost implicit in wages) do not show significance.[8] Measures that successfully predict criminal behavior speak directly to patience, self-control and mental acuity - for example, smoking or drinking behavior (negatively correlated with threshold), test scores (positively correlated), and education by age

[7]Specifically the negative binomial with mean dispersion (sometimes called the negative binomial 2 or NB2) and the Poisson.

[8]There is some evidence that direct opportunity cost is difficult to define and measure in adolescence, since overall human capital investment is much more important to lifetime income than realized wages in this period. See Lochner (2004).

16-17 (positively correlated). Given the very short nature of the average theft career, theft appears to be substantially a phenomenon of high impatience individuals entering a period of intensified risk-taking in adolescence.

Using covariates that measure impatience and asociality, I am able to explain a relatively high proportion of variation - that is to say, some of the binary discrete models have pseudo-R^2's of roughly 10%. Obviously, a significant fraction of variation remains unexplained and the estimates using count data argue strongly that there is unobserved heterogeneity. Given the stylized facts of the literature, one of the most plausible sources of such heterogeneity is social and peer effects. The NLSY data provides a few measures of peer activity in the first round, but not thereafter, so my ability to investigate social norms with this data is limited. I discuss the potential for future research in this area in the conclusion.

The major contribution of this paper is to propose, test and estimate a model of theft that parsimoniously explains a number of patterns in the NLSY and other data sets. Specifically it provides a simple explanation of the wide variation in the number of theft acts and earnings per act of theft, the instability in per-act earnings, the frequency of thieves reporting zero per-act earnings, the role of impatience and poor judgment in predicting theft activity, and the negative correlation between number of acts and per-act earnings. Moreover the structure provided by the model and its estimation, combined with the novel finding of criminal behavior spikes in adolescence, highlights the relative promise of several areas of research. In particular, the three areas that seem most fertile for future research in criminal behavior choices are first, measuring and predicting variation in individual discount rates and intertemporal preferences; second, measuring and predicting variation in peer effects on a *short-term* basis (as phenomena that operate powerfully on a monthly basis); and third, further investigation of the behavioral issues specific to adolescence and how they link to standard economic models.

There are two other contributions of this paper.

One is the disaggregation of the adolescent increase in theft activity into short spikes of theft activity. There does not appear to be an earlier direct reference to this, although it is consistent with other data, including aggregated data on property crime and the arrest records of the famous Wolfgang study of juvenile delinquency in Philadelphia. It seems likely that more than half of all acts of theft are committed by individuals who are active for less than 2 years in adolescence.

The final contribution is to demonstrate a case where only one of the standard count data models is able to match the marginal effects of the binary discrete choice models, eliminating a number of models that might have been accepted in an investigation that was not guided by theory. This suggests that the use of binary discrete and count models in tandem allows for more powerful inference than using one or the other class in isolation.

2.2 Model

I present a simple model of theft in which all individuals draw an opportunity to commit a crime on a regular basis, roughly daily. Each individual in each period compares the expected value of the opportunity he drew with his expectation of the costs, and takes any opportunity with expected benefits greater than the expected costs.

Every period t, roughly daily, each individual i gets an opportunity to commit a crime. The expected value of the opportunity is r_{it}. These expected values are distributed lognormally, with the same moments μ, σ^2 for all individuals and all periods, and are stochastically independent over time. I use the notation $f_r(r_{it}; \mu, \sigma)$ for the density and $F_r(r_{it}; \mu, \sigma)$ for the cumulative density. The expected value r_{it} is a simple summary of the inherent value of the object and the probability that the thief will either be able to use it or resell it.

The agent compares the expected value r_{it} of the theft with his expectation of the cost. The total expected cost (in opportunity cost, expected disutility from the possibility of future punishment, etc) that each agent will incur for committing a crime is individual specific.

Given the radical shifts in criminal behavior from year to year it is necessary to allow for this threshold to move at certain points in time. For this purpose, I distinguish between multiple daily periods, indexed t, and their aggregations into time intervals of length τ, roughly a year, indexed with T. The relationship can be visualized as follows:

$$
\begin{aligned}
t \text{ (roughly daily) } &= \underbrace{1\ 2\ 3\ \ldots\ \tau}\ \underbrace{1\ 2\ 3\ \ldots\ \tau}\ \underbrace{1\ 2\ 3\ \ldots\ \tau}\ \ldots \\
T \text{ (roughly yearly) } &= \qquad 1 \qquad\qquad\quad 2 \qquad\qquad\quad 3 \qquad\quad \ldots
\end{aligned}
$$

The threshold in a particular time interval is denoted ξ_{iT}. It depends on some vector X_{iT} of individual characteristics including but not limited to wage w_{iT}, age a_{iT}, mental health m_{iT}, etc. and we can write $X_{iT} = (w_{iT}, a_{iT}, m_{iT}, \ldots)$, so $\xi_{iT} = \xi(X_{iT}, \varepsilon_{iT})$. For simplicity it is assumed that $\varepsilon_{iT} \sim N(0, \sigma_\varepsilon)$ is constant within a time interval.

Thus theft opportunities are highly stochastic within a short period, while the costs are stable. The random nature of theft opportunities seems consistent with an activity that depends on momentary or minimal lapses in security and monitoring (to some extent the classic "$20 bill on the sidewalk" logic applies here - if there is a consistent reliable opportunity to steal a set amount of money on a periodic basis, than some property owner(s) are being shockingly careless). The data itself (as discussed below) support this interpretation, with highly variable mean earnings for thieves across periods. Implicitly there is some grouping here - by definition costs in instantaneous risk of detection are being set aside, and the costs I focus on, such as opportunity costs, fear of punishment, etc. are likely to be fairly stable for an individual.

Although the intervals help to explain the shifts in theft behavior from year to year, and prepare the way for future work with the panel data, the analysis presented in this paper works only in cross-section. The estimation in this paper focuses on comparing specifications across a number of nonlinear functional forms, several of which are limited or unavailable in panel form. Preliminary investigation shows that the patterns in cross-section carry over to binary discrete panel estimation (there is a brief discussion in section 2.6).

The specific functional form relating ξ_{iT} to X_{iT} and ε_{iT} is

$$\xi_{iT} = e^{X_{iT}\beta+\varepsilon_{iT}} = e^{X_{iT}\beta}e^{\varepsilon_{iT}}$$

We get the basic model

$$c_{it} = \begin{cases} 1 \quad \text{if} \quad r_{it} \geq \xi_{iT} = e^{X_{iT}\beta+\varepsilon_{iT}} \\ 0 \quad \text{else} \end{cases} \tag{2.1}$$

The probability of a single draw in a single period t leading to a theft is thus:

$$P(c_{it} = 1) = \Pr(r_{it} \geq \xi_{iT}) = 1 - F_r(\xi_{iT}|\mu, \sigma) = \int_{\xi_{iT}}^{\infty} f_r(r|\mu, \sigma)dr \tag{2.2}$$

In plain words, the indicator function $c_{it}(r_{it}, \xi_{iT})$ tells us the crime is committed if the benefit outweighs the cost, and not otherwise.

For the purposes of the analysis that follows it is convenient to take logs of both sides and re-express the model as:

$$c_{it} = \begin{cases} 1 \quad \text{if} \quad ln(r_{it}) \geq ln(\xi_{iT}) = X_i\beta + \varepsilon_{iT} \\ 0 \quad \text{else} \end{cases} \tag{2.3}$$

It is obvious that both $ln(r_{it})$ and ε_{iT} are normally distributed, and thus equation 2.2 can be re-expressed as

$$P(c_{it} = 1) = \Pr\left(ln(r_{it}) \geq X_{iT}\beta + \varepsilon_{iT}\right) = 1 - \Phi\left(\frac{(X_{iT}\beta + \varepsilon_{iT}) - \mu}{\sigma}\right) \tag{2.4}$$

The model is exactly the same whether expressed in equations 2.1 or 2.3, and the probabilities are exactly the same whether expressed in equations 2.2 or 2.4. The intuition of the model is most easily grasped using the forms of 2.1 and 2.2, while many of the computations are most easily made in the forms of 2.3 and 2.4. Throughout the remainder of the paper I use the version that is most expedient for the immediate purpose.

While the exact moments μ, σ are left undetermined for the moment, I add the following assumption.

Assumption 1. *If $\underline{\xi_{iT}}$ is the lowest value of ξ_{iT} we observe among all I individuals across all time intervals $\bar{T}$, then $f_r(r;\mu,\sigma)$ is decreasing in r_{iT} for all $r_{iT} \geq \underline{\xi_{iT}}$.*

That is, at no point in the range of opportunities large enough to interest an individual are there two points, r_a, r_b such that $r_a > r_b$ and $f_r(r_a;\mu,\sigma) > f_r(r_b;\mu,\sigma)$.

This model is given a simple graphical representation in figure 2.1. The curve shows the distribution of crime opportunities. It's highly likely that the best opportunity in a period will not be particularly promising - shoplifting a good that the respondent doesn't need or would have difficulty reselling. This is represented by the area under the curve but close to the y axis - high probability, low value opportunities. However, higher value, low probability opportunities will also come along - the individual might walk down the street and find a Bentley, unattended, with the keys in the ignition and nobody around. These are represented by the area under the curve, far out to the right. Each agent has a cutoff (the righthand side of our inequality, ξ_{iT}), based on his cost of committing the crime. Each agent, in each period t, makes a very simple calculation - if the opportunity he has drawn in the period is greater than the cutoff, he takes it. If not, he lets it slide by.

2.2.1 The Model as a Data-Generating Process

I have described the model on a period-by-period, daily basis, but the data available is aggregated at the annual level and thus it is impossible to observe c_{it} or r_{it} directly. In addition, because of the limitations of some of the estimation procedures I focus on a cross-sectional analysis that aggregates across a full seven years. Aggregating in this way across multiple time intervals there are four variables available in the data:

$$C_{iT} = \max_{t=\{1,\ldots,\tau\}} c_{it} \quad \text{for individual } i$$

$$k_{iT} = \sum_{t=1}^{\tau} c_{it} \quad \text{for individual } i$$

$$R_{iT} = \sum_{t=1}^{\tau} r_{it}c_{it} \quad \text{for individual } i$$

$$\bar{r}_{iT}|c_{it} = 1 = \frac{\sum_{t=1}^{\tau} r_{it}c_{it}}{k_{iT}} \quad \text{for individual } i$$

That is, looking at data that aggregates across a year or more we can observe whether or not an individual has stolen during a time interval (C_{iT}), we can observe the number of times the individual stole over a time interval (k_{iT}), we can observe total earnings from theft (R_{iT})and we can observe the average take per act of theft ($\bar{r}_{iT}|c_{it} = 1$).

In order to meaningfully analyze these four variables it is important to discuss their asymptotic behavior under the model. This asymptotic behavior is largely straightforward, but is somewhat complicated by the fact that for an individual i the sequence c_{it} (that is, the indicator for whether or not a theft takes place in a particular period), and hence also $r_{it}c_{it}$ (the earnings from an act of theft) are not iid but instead are stationary and exchangeable sequences. In nearly every case this leads to the same conclusion as would obtain with an iid sequence, but in each case requires a slightly more circuitous route.

The sequences of c_{it} and $r_{it}c_{it}$ are not independent because of the random error term ε_{iT} that determines ξ_{iT} in any time interval. Thus, if we think of the draw ε_{iT} occurring at the beginning of the interval (January 1st, say, at 12:01 AM), every member of the sequence

c_{it}, $t \in \{0, 1, \ldots, \tau\}$ is conditioned on that realization of ε_{iT}. They are *conditionally independent* and hence exchangeable and stationary, but not *independent*, which subtly alters our ability to determine their asymptotic behavior.

The Indicator C_{iT}

Given the model, particularly equation 2.3 it is clear that

$$C_{iT} = \begin{cases} 1 & \text{if} \quad \max_{t\in\{1,\ldots,\tau\}} \; ln(r_{it}) \;\; \geq \;\; X_{iT}\beta + \varepsilon_{iT} \\ 0 & \text{else} \end{cases}$$

which can be reexpressed

$$C_{iT} = \begin{cases} 1 & \text{if} \quad \rho_{iT} = \max_{t\in\{1,\ldots,\tau\}} \left(ln(r_{it}) - \varepsilon_{iT}\right) \;\; \geq \;\; X_{iT}\beta \\ 0 & \text{else} \end{cases}$$

Since both $ln(r_{it})$ and ε_{iT} are normally distributed, their difference is normally distributed as well. The sequence of their difference is stationary, and has a conditional covariance of zero. Berman (1964) shows that the maximum value of ρ_{iT} over τ periods as τ goes to infinity should thus converge to a Gumbel distribution. Generally for convergence to the Gumbel, by $\tau > 100$ "the agreement is already good" (David (1981), page 264). Note that because estimation results below, for the Gumbel and other models, are in cross-section not panel, technically I will be using data that measures individual traits across 1997 to 2003; the results should be thought of as for the minimum X_{iT} across the period 1997-2003.

The Count k_{iT}

It can be immediately seen that across τ time periods k_{iT} is distributed binomially as

$$k_{iT} \sim \binom{\tau}{k} \left(\int_{\xi_{iT}}^{\infty} f(r_{it}; \mu, \sigma) dr_{it} \right)^{k} \left(1 - \int_{\xi_{iT}}^{\infty} f(r_{it}; \mu, \sigma) dr_{it} \right)^{\tau - k}$$

with expectation

$$E(k_{iT}) = \tau(1 - F_r(\xi_{iT}|\mu, \sigma)) \tag{2.5}$$

That is, the expectation is the total number of draws multiplied by the probability of an act of theft in each period. As discussed above, the sequence c_{it} is not iid but is exchangeable. If we normalize the sequence by subtracting from each member the conditional expectation ($\Phi(\xi_{iT})$) and dividing by the conditional variance ($\Phi(\xi_{iT})(1-\Phi(\xi_{iT}))$) than we can apply the CLT for exchangeable sequences (summarized by DasGupta (2008)). This in turn means that across τ draws the expectation will show a binomial distribution, and we can approximate it using standard count distributions such as the Poisson or negative binomial for τ larger than 100.

Total Earnings from Theft R_{iT}

Under the model R_{iT} is the sum of τ realizations of $r_{it}c_{it}$. While we are not able to observe τ directly, we can estimate it a number of ways, and thus can use it as a parameter. This means in turn that we can generate estimates of $\frac{R_{iT}}{\tau}$ from the data. As with c_{iT} this is not an iid sequence, but is exchangeable, and under the Central Limit Theorem for exchangeable sequences will approximate a normal distribution around the mean of the lognormal distribution of theft opportunities above the point of truncation ξ_{iT}:

$$\frac{R_{iT}}{\tau} \sim N\left(E(r_{it}c_{it}), \frac{Var(r_{it}c_{it})}{\tau}\right)$$

In Aitchison and Brown (1969) we find that the formula for the expected return from crime in period t for agent i is

$$E(r_{it}c_{it}) = exp(\mu + \frac{1}{2}\sigma^2)(1 - F_r(\xi_{it}|\mu + \sigma^2, \sigma^2)) \tag{2.6}$$

and the variance is

$$Var(r_{it}c_{it}) = exp(2\mu + \sigma^2)(1 - F_r(\xi_{it}|\mu + 2\sigma^2, \sigma^2)) \tag{2.7}$$

The Average Realized Value Per Act of Theft $\bar{r}_{iT}$

Under the model $\bar{r}_{iT}$ is the sample mean of a sample of k_{iT} observations of the population mean of the distribution above ξ_{it}. Using Aitchison and Brown (1969, p. 87) we find that the formula for the expected return from crime in any particular period for agent i is

$$E(r_{it}c_{it}|c_{it}=1) = exp(\mu + \frac{1}{2}\sigma^2)\frac{1-F_r(\xi_{it}|\mu+\sigma^2,\sigma^2)}{1-F_r(\xi_{it}|\mu,\sigma^2)} \tag{2.8}$$

Unfortunately, because these values have stochastic parameters (i.e. ξ_{iT}) and are distributed according to a ratio without a straightforward distribution, we cannot effectively apply either a standard CLT or the exchangeable sequences CLT. There is no reason to believe that $\bar{r}_{iT}$ is "badly behaved" but because I am unable confirm its asymptotic behavior I will use it only sparingly.

However, the first three variables are fully predictable and thus under the model we have specific predictions for how the data will look, specifically the relationship between the outcome variables of choice to steal or not over a time interval, number of acts of theft over a time interval, and average return per act, and parameter values of $X_{iT}, \beta, \mu, \sigma,,$ and τ.

2.2.2 Likelihood Equations

Likelihood for C_{iT}

As discussed in section 2.2.1, under the model the decision to steal is determined by the relationship of the threshold ξ_{iT} and the maximum draw from the theft opportunity distribution $\max_t r_{it}$. As a result, the likelihood function for C_{iT} is a simple binary discrete choice model with a Gumbel distribution:

$$lnL = \sum_{i=1}^{n} ln(l_i)$$

where

$$l_i = \begin{cases} exp(-e^{-X_i\beta}) & \text{if} \quad C_{it}=1 \\ 1-exp(-e^{-X_i\beta}) & \text{if} \quad C_{it}=0 \end{cases}$$

Likelihood for (R_{iT}, k_{iT}) Pairs

Using the analysis from sections 2.2.1 and 2.2.1, we can see that the likelihood expression for an (R_{iT}, k_{iT}) pair can be constructed by combining the predicted behavior for each of the two variables.

The count k_{iT} is generated by a binomial distribution, from τ trials of p_i probability of success each, where $p_i = \Phi(\xi_{iT}) = \Phi(X_i\beta + \varepsilon_{iT})$. I use the Poisson distribution as an approximation.

The total return R_{it} is the sum of τ members of the sequence $r_{it}c_{it}$. If we divide by an estimate of τ we get the sample mean R_{it}/τ which under the CLT for exchangeable variables should converge to the mean of the truncated lognormal distribution. We will only observe this return in cases where $k > 0$, thus one of the two variables is censored.

The likelihood for the pair

$$lnL = \sum_{i=1}^{N} ln(l_i)$$

is then constructed

$$l_i = \begin{cases} \phi\left(\frac{R_i/\tau - \mu_{r_i}}{\sigma_{r_i}/\sqrt{k_i}}\right) \times exp(-\tau p_i)(\tau p_i)^{k_i}/k_i! & \text{if} \quad k_i \geq 0 \\ \\ exp(-\tau p_i) & \text{else} \quad k_i = 0 \end{cases}$$

Where

$$\begin{aligned} k_i &= \text{Number of acts of theft (observed)} \\ R_i &= \text{Total take across all periods (observed)} \\ p_i &= \text{Probability of committing theft} \\ \mu_{r_i} &= \text{The expectation of the take for individual } i \\ \sigma_{r_i} &= \text{The standard deviation of the take for individual } i \\ \tau &= \text{Total number of periods and hence opportunities} \end{aligned}$$

And the three expressions $p_i, \mu_{r_i}, \sigma^2_{r_i}$ are as follows:

$$p_i = \int_{exp(X_i\beta+\varepsilon)}^{\infty} f(r;\mu,\sigma^2)dr = \Phi\left(\frac{\mu - X_i\beta - \varepsilon}{\sigma}\right)$$

$$\mu_{r_i} = exp(\mu + \tfrac{1}{2}\sigma^2)\int_{exp(X_i\beta+\varepsilon)}^{\infty} f(r;\mu+\sigma^2,\sigma^2)dr = exp(\mu + \tfrac{1}{2}\sigma^2)\Phi\left(\tfrac{\mu+\sigma^2-X_i\beta-\varepsilon}{\sigma}\right)$$

$$\sigma^2_{r_i} = exp(2\mu+2\sigma^2)\int_{exp(X_i\beta+\varepsilon)}^{\infty} f(r;\mu+2\sigma^2,\sigma^2)dr = exp(2\mu+2\sigma^2)\Phi\left(\frac{\mu+\sigma^2-X_i\beta-\varepsilon}{\sigma}\right)$$

Notice that while the count is modeled as a Poisson process, the Poisson mean is not modeled as $exp(X_i\beta)$ as is standard, but instead as $\tau\Phi(X_i\beta+\varepsilon)$ to emulate the actual model. The use of τ as an exposure measure is unusual but not unprecedented (see the opening of Cameron and Trivedi (1998)).

2.2.3 Other Predictions

In a population of individuals $i = \{1,2,3,\ldots,N\}$, where ξ_{iT} is randomly distributed, the number of crimes an individual commits in a period, k_{iT} will negatively covary with his per-act earnings from theft, $\bar{r}_{iT}$

Proof. By inspection of equation 2.5 we can see that for identical τ, μ, σ it is the case that $E(k_{iT})$ is decreasing in ξ_{iT} and $E(r_{it})$ is increasing in ξ_{iT}.

Schmidt (2003) proves that two random variables $A(x), B(x)$ that are increasing monotone functions of a third variable x will be positively covariant. The corollary that if $B(x)$ is decreasing $cov(A(x),B(x)) < 0$ follows from the fact that $-B(x)$ is increasing and hence

$cov(A(x), -B(x)) = -cov(A(x), B(x)) > 0$[9].

We can therefore see that, since $E(k_{iT})$ is decreasing in ξ_{iT} and $E(\bar{r}_i)$ is increasing in ξ_{iT}, k_{iT} and $\bar{r}_i$ covary negatively. $\square$

It should be noted that both $E(k_{iT})$ and $E(\bar{r}_{iT})$ are increasing in μ, so in an alternative model where different individuals faced distributions that differed in mean (one possible mechanism by which there could be different abilities at theft), we would be likely to see *positive* covariance between k_{iT} and $\bar{r}_{iT}$.

2.2.4 Discussion of Model

Several points are worth strong emphasis.

First, within the model, *every* individual faces this distribution. Those who don't steal *choose* not to, based on their perception that the costs are higher than the opportunity.

Second, the assumption that all individuals face a common distribution is fairly strong, and given the noise in the data, it is difficult if not impossible to test definitively. However, there are several patterns in the data that are strongly consistent with it. First, the assumption implies that there is no such attribute as "theft ability" - that is, there is unlikely to be stability in earnings from theft, and a thief who earns a great deal one year might earn very little the next. Second, the assumption implies that regional and locational measures will show little or no impact on either earnings from theft or number of acts of theft. Third, a model with a common distribution of opportunities and variable thresholds will generate a negative correlation between number of acts and per act earnings, while, in contrast, variable opportunities and common thresholds would generate a positive correlation. In all cases the patterns in the data are consistent with the theory. More generally, while dividing covariates into predictors of expected cost of theft and predictors of expected benefits of

[9] I am grateful to Richard McLean for guidance on this issue, as well as for directing me to Schmidt (2003).

theft is an imperfect art, essentially all measures that show significance in the estimates are measures that both common sense and previous research would associate with expectations of cost, particularly impatience and overconfidence. With that said, it is more than likely that there are limits to the assumption. For example, there may be minute variation in the opportunity distribution available to individuals in the data which is not visible due to noise. Alternatively, there is some evidence that a small group, no more than 5% of all thieves or 0.5% of all respondents in the survey, are able to achieve unusually high earnings over an extended period of time (because there are too few of these individuals for reliable analysis, their reports of theft earnings are usually dropped or winsorized). Finally, assuming that the assumption holds for the United States at the turn of the millennium, it may not apply in other contexts. With these caveats, it appears that the variation in theft activity and earnings from theft can be explained using a model of a common distribution.

Third, the model ignores the fact that different types of theft possess different risks; most state criminal codes expressly divide theft into misdemeanor and felony offenses based on the value of the good stolen, whether or not force was used, and even idiosyncratic local rules (stealing produce is more seriously penalized in some states, such as California, where farming is a major industry, for example). Additionally, different types of crime trigger different levels of risk; a high profile jewel heist attracts much more police attention than shoplifting a pair of shoes. However, 95% of the thieves in the NLSY report per-act earnings less than $2,000. This is generally the lowest class of felony (called 3rd degree, Class C, Class 3, etc) in most state criminal codes. Most of the thefts thus appear to be either misdemeanor or a low level felony. The police response is unlikely to vary radically. While state criminal codes stipulate higher sentences for higher values, so that the sentence for stealing $2,000 might be twice that for stealing $1,000, the actual punishment, especially for a first offense, is likely to be very little. Lee and McCrary (2005) find that in Florida the average period of incarceration for juveniles for Index Crimes (the most serious of all crimes, including assault, homicide, dealing of hard drugs, etc.) was “from 2.7 to 6.5 weeks” and “[a]dult incarceration lengths are 3.1 to 7.4 times as long.” Thus, while the average thief might view the criminal justice punishment for stealing $100 as different from the punishment for stealing \$2,000, the 20× difference in the value of the objects stolen works out to a 2× difference in the severity of the legal offense, and all evidence suggests far less than that in actual expected punishment. So long as the disutility of punishment increases with regard to r_{it} at a slower rate than r_{it} itself for all values of r_{it}, then modeling a single threshold ξ_{iT} is correct. That is to say, if I were to complete the model by making threshold a function of expected return $\xi(r_{it})$, since $\frac{\partial \xi}{\partial r_{it}} < 1$ for all r_{it} then $r_{it} - \xi(r_{it})$ will be monotonically increasing in r_{it} and only cross the x-axis at one point, leading us back to a single threshold model.

Fourth, while the lognormal distribution of *ex ante* opportunities allows for a high degree of variation, there is likely to be even greater variation *ex post*, both in realized opportunities and realized opportunities net of punishment. The variation in realized returns will be discussed further in section 2.3.2 as part of the empirical analysis; the key issue is that even items that thieves favor as being small, light and having a high retail value (for example a cellphone or diamond ring) are not perfectly liquid. When trying to resell an illegally acquired item in a hurry, it is likely that there will be a significant loss of value. Additionally, and related to the preceding paragraph, even though the assumption of a single stable threshold of costs is reasonable, there is likely to be variation in actual punishments. This in turn will lead to variation in realized earnings net of punishment. In summary, there are sources of variation that are being left out of the model in its present form.

2.3 Data

2.3.1 Background

To test the predictions of the model I use the first seven rounds of the National Longitudinal Study of Youth, 1997 Cohort (Ohio State University. Center for Human Resource Research, for years 1997 to 2003). This data set tracks 8,984 individuals from 1997 onwards. The majority of respondents, roughly 3/4, come from a simple random sample of the United States youth population in 1997. The remaining one quarter of respondents come from an over-sampling of ethnic minority populations, specifically black and non-black hispanics. The respondents were equally ranged from the age of 12 to 16 in the initial round in 1997.

The rate of attrition is quite low both for thieves and non-thieves; retention rates in each of the seven years are over 86% for both groups, and over 88.9% for thieves.

In addition to a wide range of questions about work, earnings, education, assets, beliefs, health, family and other issues, every round of the NLSY 1997 includes a self-administered questionaire that asks respondents about potentially compromising issues such as criminal behavior. There are two sets of questions on theft in the 1997-2003 rounds: respondents are asked a number of questions about acts of theft where they stole items worth less than \$50, and a second set of questions about acts of theft where they stole items worth more than \$50. For each they are asked what methods they used to steal and (in most rounds) how many times they stole in the past year. For their theft of objects worth more than \$50 they are then asked for an estimate of the total earnings: "In [year], what was the amount of cash you received for the items you stole or would have received if you had sold them?". While many respondents admit to both types of theft, only 1,202, or roughly 12.5% of the respondents, admit to stealing items worth more than \$50 at any point. I believe that individuals who only steal items worth less than \$50 are committing petty thefts that are of minor interest at best, and can be ignored. With occasional explicit exceptions, by "thief" I am referring to a respondent who admitted to theft of at least one item worth more than \$50, and by "theft" I am referring to an act which the respondent characterized as of an item(s) worth more than \$50[10].

2.3.2 Shape

Since the NLSY data provides a measure of both the total number of acts of theft committed by each individual, as well as their reported total earnings from theft, it is straightforward

[10]In his work with the 1979 NLSY, Grogger (1998) cites the substantial evidence that self-report data on criminal activity for blacks seems to be less accurate than for whites. He finds the reported earnings data seems roughly equivalently accurate for black and white males. Given the possibility of different rates of reporting across different populations, as well as the possibility that different populations might have different patterns of behavior, I have tried when possible to perform the analysis first, for both genders and all ethnicities, secondly for males only, and finally for white males only. The final Gumbel specifications in table 2.7 follow this pattern, for example.

to divide total earnings by total number of acts to compute values for $\bar{r}_{iT}$[11]. Following the guidance of Angrist and Krueger (1999) I winsorize the earnings data of the top 95% and the resulting data is shown in Figure 2.2. Figure 2.3 shows the output from a simple simulation program based on the model, where thresholds vary but opportunities come from the identical lognormal distribution. The similarity to the data is immediately apparent.[12]

The data clearly shows a strong match with a basic prediction of the model.

The wide variation in earnings is consistent with previous studies (see Reuter et al. (1990), Wilson and Abrahamse (1992), Tremblay and Morselli (2000) and Levitt and Venkatesh (2000)).

It is also worth looking at one remarkable aspect of the data: the high percentage of thieves who report zero or close to zero per-act earnings. As can be seen from Figure 2.4, modal and median per-act earnings are quite low. Over one-quarter (303 out of 1202) of the respondents report no earnings from theft. The median earnings per act across respondents who report theft is $30.

While striking, such earnings from criminal behavior are consistent with the reports of other sources. For example, Reuter et al. (1990) find that one-quarter of active (i.e., reasonably successful and experienced) drug dealers in their sample in Washington D.C. earn only $50 per month. For a sample of thieves from the general population even lower earnings are to be expected.

The evidence suggests that even in the case of theft of items worth more than $50, the

[11]Some individuals do not report specific earnings but identify a range, i.e. earnings from theft in year T that are between $100 and $500. In computations that involve earnings I have experimented with both 25% and 50% of the range as the true value, i.e. for the range $100 to $500 I have tried both $200 or $300, with little change. Unless otherwise specified results are based on an imputation of 25% of the range.

[12]This similarity is not simply a byproduct of a threshold model, but is directly dependent on the lognormal distribution or visually identical distribution. When identical simulations are performed with the distribution of opportunities changed to uniform, not lognormal, the similarity to the data falls apart

total value of the theft is small. In a recent report on shoplifting (Doyle, 2012) it was found that the average recovery in cases where the shoplifter was arrested were just over \$100. The US Department of Justice in a survey of victims of burglary (Catalano, 2010) found that in 30% of cases less than \$250 of value was stolen, and in the majority of cases less than \$1,000 was stolen.

These values, as low as they are, reflect the value of the stolen objects to the victim, which will almost certainly be much greater than the value realized by the thief. All available information suggests it is very difficult for thieves to realize even a reasonable fraction of full value. The literature on professional fences (going back hundreds of years) is consistent on the range of prices; the best a thief can hope for from a fence is one-third of the value of the item but is more likely to be one-tenth or even less (see Walsh (1977, pp. 71-76) and Klockars (1974, pp. 113-129)). And because not all thieves will actually succeed in bringing what they steal to the fence and negotiating a deal it seems likely that in many cases thieves discard the items they steal without realizing the value[13].

2.3.3 Spikes

An important pattern in the data, previously discussed, is that theft careers are extremely spiky - generally a short burst of activity over fewer than 12 months, with no recorded activity before or after. Moreover, thieves who are active in only two or fewer years are responsible for the majority of acts and of value stolen, a pattern that does not seem to

[13] It is beyond the scope of this paper to try to match the exact pattern of returns, but an obvious extension of the model is to make the expected value of each theft opportunity a mixture of the actual value and a Bernoulli trial that determines whether the thief realizes the full value or not. To illustrate, consider a thief who views a cellphone worth \$300 lying unattended in a public space. There are two possible outcomes if he steals it. First, he may be able to sell it quickly to somebody, probably getting much less than normal market value, say \$50. Alternatively he can simply dispose of the object, getting nothing for it. Going through a specific example:

Outcome	Realized Value	Probability	Expected Value
Sell the cellphone	\$50	30%	\$15
Discard the cellphone	\$0	70%	\$0
Total			$r_{it} = \$15$

Such a mixture model would not effect the binary discrete or count results in any way. The bivariate structural model would need to be slightly adjusted to allow for more variation in returns.

have been brought out by previous researchers. There is a well-known and much analyzed pattern in the aggregate data, where delinquency peaks in the late teen years. The most famous article on this is probably Moffitt (1993), which disaggregates the activity into two groups: delinquents who are only active in adolescence (adolescent-limited) and individuals who engage in delinquent or criminal behavior for effectively their entire life (life-course-persistent). Subsequent work by Moffitt, Nagin and others (see, for example, Broidy et al. (2003), Nagin et al. (1995), and Nagin (1999)) looks at "trajectories" of behavior, particularly focusing on the life-course-persistent delinquents.

The aggregate pattern in the NLSY data is of a piece with previous studies, as can be seen in Figure 2.5, with a peak in mid-adolescence. However, if we disaggregate the data and look at the length of theft careers, we see the spikiness. Most thieves, and more importantly, the thieves responsible for most acts of theft, and most of the wealth stolen in theft, do not follow theft as a career, either part-time or full-time. Instead, most thieves show a spike in theft for a short period (the exact length of the median is unclear, but is less than 12 months) and then stop. Table 2.1 provides the details from the NLSY 1997 data.

While the pattern is particularly strong in the NLSY self-report data, it is consistent with other sources. The general pattern in aggregate data sets is that property crime behavior declines particularly rapidly from the late teen years (Farrington, 1986). Looking at the data from (Figlio et al.) broken down in a similar way in table 2.2 it can be seen that in arrests for all property crimes a large percentage, nearly the majority of arrests (42%), are of individuals who only show activity in 2 years out of the more than 2 decades that are tracked. The pattern for all crimes is not as spiky - only 30% of arrests are due to individuals active in 2 years[14].

[14]The self-report data and the arrest data are close enough to be consistent, but clearly differ. While there are a number of possible explanations that might align them, a parsimonious and reasonable explanation is that the self-report data undercount the particularly asocial chronic thieves, while the arrest reports undercount, at the same rate, the short-lived thieves. If this assumption - a kind of symmetric selection bias - is correct it would mean that in the underlying population roughly 60% of activity is due to individuals active in only two years.

2.3.4 Lack of Consistent Patterns In Earnings

Not only are theft careers generally very short, but even when thieves have long careers (that is, more than one year), they show very little consistency of earnings.

The model does not make strong predictions as to the period-to-period correlation of components of X_{iT}. Common-sense would suggest there should be some stability. The first two rows of table 2.3 show the pattern for two covariates, log of wage and self-reported health, and indeed there are decent and stable correlations throughout adolescence. Neither correlation ever drops below 15%, and both show steady increases with age.

The model does predict that the period to period correlation of earnings would be lower and less stable, and as can be seen in the last three rows of table 2.3 that is certainly the case. Correlation in log of earnings per act of theft is fairly unstable year-to-year[15], and dips close to zero around ages 18 to 20. However, this seems to be entirely due to individuals who report \$0 earnings in two years in a row. Once we limit the observations to those with positive realized per-theft earnings (in other words, once we look for "theft ability" among thieves who actually seem to earn money) even that consistency falls away, and the correlations begin to flip, apparently randomly, between negative and positive values.

The lack of correlation from year to year in per act earnings is consistent with (although it does not prove) the assumption of the model that all individuals face the same distribution of theft opportunities.

2.3.5 Inverse Correlation of Theft Career Length and Theft Earnings

Examining the scatterplot of data in Figure 2.2 we see that indeed as predicted by Theorem 1, the very active thieves with many acts appear to have extremely low earnings, while the thieves earning the most per-act steal at most only a handful of times. Computing the

[15] The correlation for earnings itself (not its logarithm) is much lower.

correlation coefficient across the annual data from 1998 to 2003[16], using only respondent-year observations where at least one theft was reported, the correlation is -0.0100, not at all significant. If we winsorize total earnings at the 95 percentile for thieves and recompute per-act earnings, we get a very strong negative correlation of -.1080 and a bootstrap standard error of .0051129, for a t-statistic of -21.11[17]. Using the cross-section data the correlation coefficient among thieves is 0.0007. If we use winsorized earnings, however (the 95 percentile value is then $10,000) the correlation goes strongly negative to -0.085148, we find a bootstrap standard error of 0.0059019, and t-statistic of -14.43.

The prediction of negative correlation from Theorem 1 is thus borne out in the data.

2.4 Structural and Semiparametric Estimation

I now review the results from maximum likelihood estimation of two structural models based on the likelihood expressions outlined in sections 2.2.2 and 2.2.2. This represents an important empirical test of the theoretical model. The first estimation is a simple binary discrete choice model estimating $X_{iT}\beta$ using our indicator variable C_{iT} of whether or not an individual has committed any crime over a time interval. The second is a bivariate semi-censored model that incorporates the count of acts of theft k_{iT} for all individuals and the total earnings from theft R_{iT} in those cases where $k_{iT} > 0$.

I use two approaches to evaluate the success of the theoretical model.

The first method is to compare marginal effects across a number of specifications, paying particular attention to how well the structural models match semiparametric estimations (the semiparametric estimation is done both for binary data and count data). By this measure, the structural models perform extremely well, matching the semiparametric estimations across both binary and count data estimations, and both sets (count and binary)

[16]For consistency reasons, dropping 1997.

[17]There are 1,213 respondent-year observations where theft takes place.

match each other.

For this first method of comparing marginal effects I use a fixed group of seven covariates: smoking behavior and alcohol consumption in the mid-teen years, ASVAB test scores, mental health evaluation in 2000, age in 1997, and dummy variables for gender and African-American ethnicity. These seven covariates can be divided into measures or correlates of patience, self-control and foresight (the first four) and simple demographic controls (the last three). Of the first four measures, all of them are consistently and strongly significant across a range of specifications, and are nearly the only covariates of which this is true. Of the three controls, only gender shows significance.

The second method for evaluating the model is to compare the significance and explanatory power of covariates (as measured by pseudo-R^2), for simplicity focusing on the binary discrete model. In this second approach I use a number of specifications (only some of which are reproduced here) to see which covariates provide significant explanatory power. As discussed above, the variables which show explanatory power are almost universally related to how an individual weights present and future consequences.

2.4.1 Estimating the Binary Discrete Choice Model

In the model section it was shown that under the assumptions of the model ρ_{iT} is the maximum value draw from the opportunity distribution, is the maximum of a normal distribution, and should thus follow a Gumbel distribution. I maximize the likelihood from section 2.2.2:

$$lnL = \sum_{i=1}^{n} ln(l_i)$$

where

$$l_i = \begin{cases} exp(-e^{-X_i\beta}) & \text{if} \quad C_{it} = 1 \\ 1 - exp(-e^{-X_i\beta}) & \text{if} \quad C_{it} = 0 \end{cases}$$

All results I get from the multiple specifications used to compare marginals are presented in tables 2.4 and 2.5 so that they can be easily compared. The β estimates for both structural models, and the three reduced form count models can be seen in table 2.4. The marginal effects for these five models and the two semiparametric estimations can be seen in table 2.5. Focusing on two continuous variables that show strong significance across specifications, I provide information on the marginal effects for smoking behavior (number of days in the past month that the respondent smoked at age 16/17) and mental health (the results on a scale of 20[18] of an objective evaluation of mental health undertaken in 2000).

Looking at the first two columns of table 2.5 we see that in both cases the marginal estimates for the semiparametric binary and the Gumbel distribution provide a clear match. The median value of the marginal for cigarette use is 0.0036 and 0.0035 respectively; that is, an increase in cigarette usage of one day per month in mid-adolescence is associated with an increase in the probability of becoming a thief of roughly 1/3 of one percent. The median value of the marginal for mental health is 0.0064 and 0.0065 respectively; that is, an increase in evaluated mental health of one unit is associated with a decrease in the odds of ever stealing of roughly 2/3rds of a percent.

2.4.2 Estimating the Model as a Bivariate Semi-Censored Distribution

The likelihood expression for the bivariate version of the structural model (that is, for (R_{iT}, k_{iT}) pairs) is given at the close of the discussion of the model above. It incorporates a normally distributed error term that sets unobserved heterogeneity.

There are a number of ways of incorporating such a term. In the case of this structural model it is impossible to integrate out such a term, and so either simulation methods (Markov Chain Monte Carlo or Maximum Simulated Likelihood) or numeric integration methods are required. I use Gauss-Hermite quadrature for simplicity (later extensions may use MCMC or MSL techniques as a further robustness check and/or as a way to adapt the structural

[18]Only 12 units in the scale are used in the observations I use.

model for panel data).

Gauss-Hermite quadrature requires a revision of the likelihood function.

$$lnL = \sum_{i=1}^{N} ln(l_i)$$

$$l_i = \begin{cases} \sum_{g=1}^{G} \quad \frac{1}{\sqrt{\pi}} w(x_g) \phi\left(\frac{R_i/\tau - \mu_{r_i}}{\sigma_{r_i}/\sqrt{k_i}}\right) \times exp(-\tau p_i)(\tau p_i)^{k_i}/k_i! & \text{if} \quad k_i \geq 0 \\ \\ \sum_{g=1}^{G} \quad \frac{1}{\sqrt{\pi}} w(x_g) exp(-\tau p_i) & \text{else} \quad k_i = 0 \end{cases}$$

Where

k_i	=	Number of acts of theft (observed)
R_i	=	Total take across all opportunities(observed)
p_i	=	Probability of committing theft
μ_{r_i}	=	The expectation of the take for individual i
σ_{r_i}	=	The standard deviation of the take for individual i
τ	=	Total number of periods and hence opportunities
G	=	Total number of Gauss-Hermite points (usually 2)
$\frac{1}{\sqrt{\pi}} w(x_g)$	=	Gauss-Hermite weights, where $w(x_g)$is the standard weight.
x_g	=	$\sigma_g \sqrt{2} a_g$ = Gauss-Hermite abscissas, where $\sigma_g$is the GH variance term and a_g is the standard abscissa.

$$p_i = \int_{exp(X_i\beta + x_g)}^{\infty} f(r; \mu, \sigma^2) dr = \Phi\left(\frac{\mu - X_i\beta - x_g}{\sigma}\right)$$

$$\mu_{r_i} = exp(\mu + \frac{1}{2}\sigma^2) \int_{exp(X_i\beta + x_g)}^{\infty} f(r; \mu + \sigma^2, \sigma^2) dr = exp(\mu + \frac{1}{2}\sigma^2) \Phi\left(\frac{\mu + \sigma^2 - X_i\beta - x_g}{\sigma}\right)$$

$$\sigma^2_{r_i} = exp(2\mu{+}2\sigma^2)\int_{exp(X_i\beta+x_g)}^{\infty} f(r;\mu+2\sigma^2,\sigma^2)dr = exp(2\mu{+}2\sigma^2)\Phi\left(\frac{\mu+\sigma^2-X_i\beta-x_g}{\sigma}\right)$$

Additionally I use semiparametric multiple discrete choice as a check[19].

In theory it would be possible to estimate the number of periods that thieves are active by leaving it as a free parameter. In practice this fails, as a very high number of periods inherently allows for a higher likelihood; with a parameter estimate of 1 million periods the difference between no acts of theft and 100 acts is very small. I have maximized the likelihood with set values of the number of periods ranging from 100 to 1,000. For high numbers of periods the standard errors become improbably small. I find the most reasonable parameterizations come when it is assumed that the average thief is active for 100-300 periods, and I use 250 as the number of periods in the estimation results reported.

The parameter estimates for the structural model are reported in column 2 of table 2.4. The parameters for the lognormal distribution are reasonable and are not far from the values used in the simulation (the estimated mean and standard deviation are -1.314 and 2.153 respectively, while the mean and standard deviation in the simulations are 0 and 1). Additionally, the Gauss Hermite quadrature shows a standard distribution of 1.7, which in the context of the parameter estimates shows that the error term ε plays a significant role.

The beta estimates are similar to those for other estimates. The marginal effects are discussed further below.

Using the parameter estimates we can generate estimates of the average thresholds, which are of interest. The mean threshold (which can be thought of as the minimum expected cost of theft the respondent holds during the seven years from 1997-2003) is \$1450. We cannot compute the actual Gauss Hermite error terms for each respondent, but looking at the effect of a one standard deviation positive shock and a one standard deviation negative

[19] For the semiparametric multinomial discrete choice the categories I divide the counts into are $\{\{0, 1, 2, 3, \{4, \ldots, 8\}, \{9, \ldots, 20\}, \{21, \ldots\}\}$

shock it appears that most respondents have thresholds in the range of \$7-\$30,000 at the period when their threshold is lowest.

Marginal effects for smoking and mental health for the semiparametric count and structural models are reported in the third and fourth columns of table 2.5, and the distribution of the smoking marginal effects can be seen in figure 2.7. For smoking, both models match the binary discrete models. For mental health the structural model matches the other estimates, while the semiparametric count provides higher marginal effect estimates. In the case of the semiparametric estimate, it's not clear what causes the increase in marginal effect but may be due to an outlier effect.

Allowing the Mean of the Distribution of Theft Opportunities to Vary

A further advantage of the structural model is it allows us to test the assumption that all thieves have access to the same distribution. In the first specification of the structural model, the distribution of opportunities is the same for all thieves (the mean of the distribution is -1.314 for *everybody*).

By allowing the mean of the opportunity distribution to vary according to individual characteristics, we can loosen this restriction. The results for one such specification can be seen in the third column, where the mean of the opportunity distribution is determined by a constant as well as effects for region the respondent lived in 1997 and whether or not the respondent was urban in 1997. The log-likelihood for this specification is -5810.829 as opposed to -5814.145. Using the likelihood ratio test we get a test statistic of

$$-2(-5814.145 - (-5810.829)) = 6.632$$

Since the assumption that the mean is constant involves four restrictions, we can find the p-value for the χ-squared distribution with 4 degrees of freedom is 84%, thus not significant and further vindicating the key assumption of the model that all thieves (and non-thieves) have access to the same distribution of opportunities.

This pattern holds across a range of different specifications; allowing the mean to vary across individuals does not significantly improve the log-likelihood.

2.5 Reduced Form Estimation

2.5.1 Count Models

As a test of the basic relationships of the model I run a reduced form specification of the basic model using the standard count regression procedures: Poisson, negative binomial with constant dispersion (NB1) and negative binomial with mean dispersion (NB2). Some discussion of the differences between the NB1 and the NB2 is in order[20].

Comparison of the Negative Binomial 1 (Constant Dispersion) and Negative Binomial 2 (Mean Dispersion)

The variable being regressed is k_i, the count of acts for individual i (ignoring time subscripting). The core of each regression equation is the Poisson model, with $\lambda_i = exp(X_i\beta)$ as the simple Poisson parameter of individual i. Because there is no unobserved error term under the Poisson parameterization λ_i is both the expectation of k_i and the variance. (Notice that for the reduced form specification the Poisson parameter does not utilize a normal cumulative probability p_i nor an exposure term τ as the structural model does.)

Both NB1 and NB2 allow for overdispersion by adding a gamma-distributed error term ν_i to the Poisson parameter:

$$\lambda_i^* = exp(X_i\beta + \nu_i)$$

The NB2 parameterization is

$$\lambda_i^* \sim \text{Gamma}(1/\alpha, \alpha\lambda_i)$$

[20]The discussion below is fairly standard and borrows from Cameron and Trivedi (1998), Greene (2007) and Stata Corporation (2009).

$$Var(k_i) = \lambda_i + \alpha\lambda_i^2$$

where α is a strictly positive parameter. Notice that the variance of k_i increases as a positive quadratic function of the parameter λ_i (the estimate of the expected count for individual i).

The NB1 parameterization is

$$\lambda_i^* \sim \text{Gamma}(\lambda_i/\delta, \delta)$$

$$Var(k_i) = \lambda_i(1 + \delta)$$

where δ is the parameter for the gamma distribution. Notice that the variance of the NB1 only increases linearly in the expected count.

How do the two negative binomial models relate to the structural model? The structural model is a mixture, where the random variable ε determines k_i conditional on $X_i\beta$. Using the conditional variance formula I can write

$$Var(k_i) = E(Var(k_i)|\varepsilon) + Var(E(k_i)|\varepsilon)$$

Since in the model k_i is being approximated with a Poisson process, the variance and the expectation of k_i are simply τp_i, where $p_i = \Phi(\xi_i) = \Phi(X_i\beta + \varepsilon)$ this can be rewritten:

$$Var(k_i) = \underbrace{\tau\ E(\Phi(X_i\beta + \varepsilon))}_{First\ \ Term} + \underbrace{Var(\tau\Phi(X_i\beta + \varepsilon))}_{Second\ \ Term}$$

The first time is clearly linear. Using the delta method we can say that the second term is approximated by the expression

$$\tau^2\sigma^2\phi^2(X_i\beta + \varepsilon)$$

The expression $\tau^2\sigma^2$ is a constant where σ is the variance of the error term ε, and the expression $\phi^2(X_i\beta + \varepsilon)$ varies from 0 to $\frac{1}{2\pi}$ in value, with the lowest values when $X_i\beta + \varepsilon$ is at its highest.

The variance of k_i under my model is thus stable or decreasing in the square of $X_i\beta$, and thus an NB1 parameterization is to be expected to be a better fit than an NB2. Partly this springs from the fact that the k_i's I am estimating represent single, relatively homogeneous individuals (humans), with a single Bernoulli trial per period. A natural contrast would be if there was variation by individuals *both* in the number of periods *and* in the per-period probability, in which case there would be the potential for the variance to increase quadratically with the mean.

Count Estimates

Output from regressions of the three standard count models (NB1, NB2 and Poisson) can be seen in table 2.4 and marginal effects can be seen in table 2.5[21]. The coefficient estimates from three models seem similar. The Poisson, which does not model unobserved heterogeneity, shows extremely high t-statistics, usually a sign that there is unobserved heterogeneity that the Poisson cannot model (Cameron and Trivedi, 1998). Additionally, the χ^2 test of the Gamma distributions that parameterize both the NB1 and the NB2 are extremely high - over 23,000, yielding an astronomically low p-value - all of which is strong evidence of significant unobserved heterogeneity. The Poisson model can thus be rejected. The log-likelihood for the NB2 is significantly lower than for the NB1, but since they are not nested models this is not definitive. Overall the count model estimates further substantiate the structural model estimates of the Gauss Hermite error term that showed ε playing a significant role.

Looking at the marginal estimates in table 2.5 we see two quite remarkable results. The first

[21]The density for the Poisson is computed as

$$Pr(Y = 0|\lambda_i) = exp(-\lambda_i)$$

where $\lambda_i = exp(x_i\beta)$. For the NB1 computation the density (from Cameron and Trivedi (1986)) is

$$Pr(Y = 0|\lambda_i, \delta) = ((\lambda_i/\delta)/(\lambda_i/\delta + \lambda_i))^{(\lambda_i/\delta)}$$

where δ is the dispersion parameter. For the NB2 computation the density is

$$Pr(Y = 0|\lambda_i, \alpha) = (\alpha^{-1}/(\alpha^{-1} + \lambda_i))^{\alpha^{-1}}$$

where α is the dispersion parameter.

is that the marginal estimates for the NB1 are virtually identical to the marginal estimates of the Gumbel, not simply at the median but across all deciles, for both smoking and for mental health, and as a result very close to the semiparametric binary estimate and close to the structural model. The second is the complete failure of the NB2. The median estimates of the marginal effects are at least somewhat off (0.0020 for smoking, versus the range of 0.0032 to 0.0040 of the other models), but the real failure is with regard to the dispersion of the marginal effects: the standard deviation of the NB2 marginal effects is less than 1/10th that of the other estimates. Essentially the NB2 has modeled variation as almost entirely due to the error term; the marginal effects are roughly identical across all individuals.

The count regression literature does not document cases where the NB1 and NB2 perform radically differently on marginal estimates. Measures of performance are usually based on tests of how the models handle overdispersion (Cameron and Trivedi, 1998). Greene (2007) provides an overarching model, the NBP, that nests both the NB1 and NB2. [22] Based on the theoretical analysis in this section, the success of the NB1 and the failure of the NB2 is further substantiation of the model.

2.6 Gumbel Models of the Decision to Steal

To explore the independent variables that determine the decision to steal, I include the results of a series of Gumbel models in tables 2.6 and 2.7.

As background, it should be noted that predicting criminal behavior using observable traits is notoriously difficult. To select a random example of work by researchers who were brave enough to reveal their R^2 values, Levitt and Lochner (2001) use data from the NLSY 1979 to predict property crime and violent crime in a single year.[23] They show two specifications for their linear probability model for property crime, the first using largely exogenous and

[22] Using the NBP with and without restrictions, the NBP's specification was not significantly different, under the likelihood ratio test, than the NB1, but both the NBP and the NB1 were significantly better than the NB2

[23] Because the NLSY 1979 only asked questions about criminal activity in a single year.

common (in the data) variables, and the second using potentially endogenous variables and variables limited to a subset (Table 7.5 in Levitt and Lochner (2001)). The first specification yields an R^2 of 0.0252, while the second yields an R^2 of 0.0523.

In table 2.6, I begin with the simple regression of the dependent dummy variable, "Did the respondent report committing theft between 1997 and 2003?" on a single independent variable, "Number of days per month that the respondent smoked cigarettes at the age of 16/17.[24]" As can be seen, the pseudo-R^2 on this regression is 0.046. There are a number of reasons this single measure is able to yield a (relatively) high explanatory power. First, by aggregating across most of the period of adolescence I am able to smooth out the spikiness discussed above. Secondly, by using a variable that measures impatience (DellaVigna and Paserman, 2005) I get to the question of how the respondents evaluate short-term benefits versus long-term costs. Finally, smoking behavior at the turn of the millennium is likely to also be indicative of a general view of social norms and of social identification.

I continue by including basic demographic variables (gender, ethnicity, age), as well as measures of family background and relationship with parents, and intellectual ability and mental health. By combining these variables and using data that aggregates across 1997-2003 I am able to get a pseudo-R^2 of 0.096, substantially above and in fact nearly double the R^2 of Levitt and Lochner (2001) (the true R^2's of similarly specified linear probability models are generally very close).

The variables that show the most predictive power are consistent with the model, in that they speak to the respondents' future costs of crime. In particular, measures that are related to foresight, patience, self-control and equilibrium - mental health, ability at intellectually demanding tasks, caution about being arrested, and avoiding risky behavior such as cigarette consumption - are all negatively correlated with the decision to steal. The pattern is strong but there are ambiguous aspects. For example, some of the covariates could

[24]Averaged in the cases where we have two observations.

enter in several ways - for example, ASVAB scores could measure both discount rate and long-term earning potential (and hence opportunity cost in lost future earnings).

As discussed in section 3.3, work with binary discrete panel regressions using fixed effects show very similar patterns as in cross-section. The only major change that appears is regional dummies; the states West of the Mississippi appear to be slightly more honest, such that residence reduces the annual probability of theft activity by 1%.

2.7 Data Robustness Checks

An obvious question with any data on crime, and self-report data in particular, is how reliable it is, and how much measurement error may be present. I have performed a number checks on this data, and find it to be largely reliable, especially for males.

The results of the first analysis, comparing rates of arrests by age in the self-report data and in the FBI Uniform Crime Reports, can be seen in figures 2.8. The rates from the data are the weighted number of charges for burglary, theft or robbery per person for the relevant ages. The rates from the FBI Uniform Crime Reports are the number of arrests for property crime for each age, divided by an estimate of the age cohort from US Census estimates[25]. The rates for all are fairly close and in fact the accuracy of the male self-reports (not shown) are even closer. It should also be said that the UCRs, while the most reliable resource for aggregate data on arrests in the United States, are not perfect or comprehensive, and that a discrepancy between the NLSY data and the UCRs is not necessarily a sign of inaccuracy in the NLSY.

A second check compares the number of respondents who report committing theft and who report being arrested for theft with the number of respondents who admit being arrested for theft but deny ever stealing; this provides some evidence about the reliability of

[25]Specifically, the denominator in the rate calculation is one-fifth of the US Census's estimate of the size of the 5-year age cohort for the middle of the relevant year, scaled to account for the fact that the UCR only cover 2/3rds of the United States population

the self-report data as well as how it compares with police records, the major alternative. The odds of a respondent who self-reports committing theft also reporting being charged with theft are about 10% (slightly more if the respondent is in his 20's), which suggests that the self-report data is significantly more reliable than police records.

There are also a number of respondents who report being charged with theft but deny committing theft. If we assume that all individuals who are charged with theft admit being charged, and that police are equally like to charge a thief who does not self-report to the NLSY then this gives us an upper bound on the number of people who might be committing theft but not self-reporting. Dividing the number of respondents who admit being charged but deny stealing by the percent of self-reporting thiefs who are charged with theft suggests that the self-report data include at least 65-70% of the individuals who are stealing.

The third check attempts to understand how those who admit to both theft and being charged differ from those who admit only to being charged. The two groups appear well-matched on some basic demographic information such as age and gender, but the proportion of non-whites is higher among those reporting arrests but not acts of theft. The most significant differences are that "honest" self-reports (who admit to both theft and charges) appear to have higher test scores but worse mental health. This suggests that there may be limited selection bias in working with the self-report data.

In summary, the data is not perfect, but seems to be roughly 65-70% accurate, especially when we focus on males. There do not appear to be a missing group of particularly capable criminals (which might challenge the assumption of a common distribution of theft opportunities) but it is possible that we are losing information on life-course persistent criminals.

2.8 Discussion and Implications

The analysis in this paper outlines a number of empirical reasons for viewing theft as in large part a short-lived phenomenon of adolescence. While the data suggests that thieves are more impatient and have less foresight than others, theft careers are so short on average (probably 6-9 months) that there must be temporary factors that trigger the beginning and end of the theft career. Obvious candidates are peer influences of some kind, or some kind of developmental stage with perhaps a biological basis.

2.8.1 Policy

The role of impatience and poor foresight in predicting thresholds and theft behavior reinforce other work, such as Lee and McCrary (2005) and Piehl and Williams (2011), suggesting there are serious limits to deterrence. Additionally, the novel finding of the spiky nature of theft activity implies that by the time a thief is caught he is likely to be close to the end of his career and the incapacitation effect of prison is unlikely to make much difference. If deterrence and incapacitation have such limits, what policy levers are left? While it does not directly support them, the analysis is consistent with the policy suggestions of Marcus Felson and others (Felson, 2009) that we may be able to actively reduce available opportunities.

Over the longer term, the work of James Heckman and others (Cunha et al., 2006) implies that early childhood interventions are likely to help increase the foresight and patience of individuals over their entire lives, thus reducing the probability of a theft career. It should also be pointed out that if pre-Kindergarten interventions are helpful, than prenatal interventions are even better. (Moffitt, 1993) speculates that a cause of life-course persistent delinquency may be damage *in utero*. There is evidence that prenatal and neonatal health interventions can have an important impact on physical and cognitive development (Deaton, 2007), and arguably general improvement in early childhood health could possibly improve foresight, temperament and self-control .

2.8.2 Adolescence, Spikes And Peer Effects

It is striking that the spikes in theft behavior take place in adolescence, when individuals are moving from being juveniles, with social circles defined by their families of origin, to adults, with social circles that they define and generate. Adolescence shows a wide range of behavioral and physiological changes, including a substantial increase in risk-taking. At the same time, there is strong evidence of a change in social links; researchers estimate that adolescents spend 1/3rd of their normal waking hours speaking with peers, and only 8% speaking with adults, and place a very high value on peer relationships (Spear, 2000). These new social links also seem to play a key role in risk-taking and there is some experimental evidence that adolescents on their own are only slightly more risk-taking than adults on their own, but are radically more risk-taking when in a group with their peers.

2.8.3 Future Research

Additional work to develop the estimators here for use with panel data is an obvious extension, to use the NLSY data fully. Moving beyond the NLSY data set, there are a number of areas that this analysis points to as fruitful for future research. First, it would be extremely interesting to learn more about the period of the spikes. Are there any visible markers that delimit it? Is the cause internal (i.e., hormonal) or external (peer influence)? Are other behaviors such as violence more prevalent during this period? Second, given the limited explanatory power of a large number of obvious variables in the NLSY data, it would be very interesting to explore two areas that the NLSY data covers only sparsely: month-to-month peer influences and hormonal/developmental changes during adolescence. Finally, the general role of discount rates and self-control in decision-making and in particular on decision-making during adolescence.

Activity in:	No. Resps.	% of total	No. of acts	% of acts	Total wealth stolen	% of wealth
1 YR	888	74%	5,484	42%	$1,877,512	28%
2 YRS	217	18%	3,853	30%	$1,985,125	29%
3 YRS	62	5%	1,615	12%	$1,301,365	19%
4 YRS+	35	3%	2,063	16%	$1,594,910	24%
	1202	100%	13,015	100%	$6,758,912	100%

Table 2.1: A table showing the spiky nature of theft careers. In all three categories: total number of individuals, total number of acts, total amount of wealth stolen, the preponderance of theft is the result of individuals who are active for a very short period; that is, by individuals who show a short, temporary, spike in theft.

Activity in	No. Individuals	% of Individuals	No. Arrests	% of Arrests
1 YR	1,862	55%	2,068	23%
2 YRS	672	20%	1,665	19%
3 YRS	343	10%	1,383	15%
4 YRS +	467	14%	3,876	43%
	3,344	100%	8,992	100%

Table 2.2: A table showing a breakdown of property crime activities in the Wolfgang data. In the two categories of total number of individuals and total number of arrests, the preponderance of theft is the result of individuals who are active as thieves (I am only looking at theft in this analysis) in fewer than 2 or three years out of 18 that are tracked; that is, by individuals who show a short, temporary, spike in theft. Given that these are for full arrests, that is for particularly noteworthy crimes, the data are consistent with the NLSY 1997 self-reports. (Property crimes are defined as crimes where the first offense per arrest is given a code from 300 to 399 or 500 to 799.)

Year-to-year correlations in the NLSY data, by respondent age
(Standard errors in parentheses, computed via bootstrap, 200 replications)

	Respondent Age						
	15 to 16	16 to 17	17 to 18	18 to 19	19 to 20	20 to 21	21 to 22
	Self-reported health						
Correlation	0.43	0.45	0.47	0.49	0.5	0.49	0.52
(SE)	(.023)	(.016)	(.012)	(.01)	(.01)	(.0095)	(.011)
95% CI: Lower bound	0.39	0.42	0.45	0.47	0.48	0.47	0.49
Upper bound	0.48	0.48	0.49	0.51	0.52	0.5	0.54
N	1634	3113	4811	6300	7465	6046	4454
	Log of wage						
Correlation	0.29	0.2	0.25	0.25	0.25	0.38	0.42
(SE)	(.11)	(.055)	(.04)	(.03)	(.026)	(.033)	(.035)
95% CI: Lower bound	0.081	0.089	0.17	0.2	0.2	0.32	0.35
Upper bound	0.5	0.3	0.33	0.31	0.3	0.45	0.49
N	168	785	2347	3993	5448	4649	3434
	Log of per-act earnings (all thieves, including estimated earnings)						
Correlation	0.16	0.2	0.045	0.14	-0.0056	0.23	0.25
(SE)	(.35)	(.24)	(.2)	(.16)	(.25)	(.24)	(.37)
95% CI: Lower bound	-0.52	-0.26	-0.34	-0.18	-0.5	-0.24	-0.49
Upper bound	0.84	0.67	0.43	0.46	0.49	0.71	0.98
N	13	29	29	45	32	20	17
	(Only thieves earning > $0 this year, including estimated earnings)						
Correlation	-0.42	0.25	-0.18	-0.15	-0.31	0.75	-0.63
(SE)	(.57)	(.11)	(.16)	(.16)	(.22)	(.11)	(.32)
95% CI: Lower bound	-1.5	0.034	-0.5	-0.46	-0.74	0.53	-1.3
Upper bound	0.71	0.46	0.14	0.16	0.11	0.98	0.0026
N	10	26	21	31	30	17	14
	(Only thieves earning > $0 both years, including estimated earnings)						
Correlation	0.82	0.26	-0.16	0.076	0.54	0.66	-0.27
(SE)	(.099)	(.2)	(.26)	(.23)	(.19)	(.16)	(.32)
95% CI: Lower bound	0.62	-0.14	-0.68	-0.38	0.17	0.35	-0.9
Upper bound	1.0	0.66	0.35	0.53	0.9	0.97	0.37
N	8	23	18	25	24	16	13

Table 2.3: A figure showing correlations from year-to-year, for ages 15 to 23, for log earnings from theft and self-reported general health and log wages. Notice that correlations for health and wages remain positive and move up as respondents move into their twenties, as would be expected of a stable trait. In comparison, per-act earnings show little to no stability in any form; notice the fluctuation of the point estimates relative to the magnitude of the standard errors.

Parameter Estimates
Dependent Variables: Respondent Reports Stealing > $50 between 1997-2003, Number of Acts, Total Theft Earnings

	Binary	Bivariate		Count		
	Gumbel	Structural Model		Negative Binomial 1 (Linear)	Negative Binomial 2 (Quadratic)	Poisson
Variable:	(Reports Stealing)	(No. of Acts and Theft Earnings)		(No. of Acts)	(No. of Acts)	(No. of Acts)
Days smoking in	-0.035***	-0.042***	-0.047***	0.035***	0.046***	0.042***
last month (age 16/17)	(-10.914)	(-18.959)	(-17.526)	(10.995)	(7.030)	(40.073)
Days consumed alcohol in	-0.020**	-0.044***	-0.034***	0.021**	0.038*	0.016***
last month (age 16/17)	(-2.936)	(-9.993)	(-6.568)	(3.050)	(2.094)	(7.242)
ASVAB Math/Verbal	0.008***	0.010***	0.008***	-0.008***	-0.010***	-0.012***
(1000 pts)	(5.030)	(10.053)	(7.605)	(-5.262)	(-3.773)	(-22.941)
Mental Health	0.067***	0.044***	0.071***	-0.071***	-0.129***	-0.093***
(eval 2000)	(4.500)	(5.000)	(7.697)	(-4.803)	(-4.935)	(-19.800)
Age in 1997	0.016	0.073***	0.067***	-0.013	-0.065	-0.004
	(0.552)	(3.983)	(3.417)	(-0.453)	(-1.282)	(-0.411)
Female	0.841***	0.918***	1.070***	-0.847***	-1.348***	-1.233***
	(9.670)	(15.568)	(17.286)	(-9.778)	(-9.530)	(-37.880)
Black	-0.039	-0.098	-0.010	0.052	-0.099	-0.102**
	(-0.383)	(-1.559)	(-0.144)	(0.519)	(-0.572)	(-2.984)
Constant	0.361	4.700***	4.474***	1.711***	3.369***	2.123***
	(0.734)	(14.884)	(14.145)	(3.480)	(3.869)	(12.946)
Lognormal Distribution Parameters						
Mean μ		-1.314***	-1.209***			
		(-7.067)	(-6.297)			
North Central in 1997			-0.148**			
			(-2.581)			
South in 1997			0.053			
			(0.853)			
Northeast in 1997			-0.068			
			(-0.956)			
Urban in 1997			0.092*			
			(2.149)			
Standard Deviation σ		2.153***	2.131***			
		(47.365)	(46.630)			
Gauss Hermite Parameter						
Standard Deviation σ_g		1.714***	1.730***			
		(40.034)	(37.791)			
Negative Binomial Dispersion Parameters						
NB1 $ln(\delta)$			3.250***			
			(38.605)			
NB2 $ln(\alpha)$				2.926***		
				(59.344)		
Log Likelihood	-1716.034	-5814.145	-5810.829	-3448.097	-3511.441	-1.53e+04
χ^2 test of dispersion				23,777	23,650	
N	4782	4782	4782	4782	4782	4782

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
t-statistics in parentheses

Table 2.4: Parameter estimates from binary, count and structural (bivariate) regressions. For the structural model 9 Gauss-Hermite points are used, and 250 periods.

Marginal Effects for Discrete Choice and Count Models
Effect: Net change in probability of ever stealing in response to a one-unit increase in independent variable.

		Binary		Count				
	Decile	Semiparametric Binary	Gumbel	Semiparametric Count	Structural Model	NB1 (Linear)	NB2 (Quadratic)	Poisson
Days smoking in last month	*10*	0.0009	0.0017	0.0007	0.0022	0.0016	0.0018	0.006
	20	0.002	0.0021	0.0014	0.0027	0.002	0.0019	0.0078
	30	0.0024	0.0025	0.0022	0.0031	0.0024	0.0019	0.0095
	40	0.003	0.003	0.0028	0.0038	0.0029	0.0019	0.011
Median:	***50***	**0.0036**	**0.0035**	**0.0032**	**0.0040**	**0.0034**	**0.0020**	**0.0124**
	60	0.0042	0.0041	0.0037	0.0055	0.0041	0.0020	0.0134
	70	0.0046	0.0048	0.0048	0.0065	0.0047	0.0020	0.0143
	80	0.0061	0.0057	0.0059	0.0076	0.0056	0.0020	0.0148
	90	0.01	0.0076	0.0089	0.0083	0.0076	0.0020	0.0152
Mean:		0.0042	0.0041	0.0038	0.0044	0.0041	0.0020	0.0111
(SE):		(0.000228)	(0.000395)	(0.000282)	(0.001036)	(0.000359)	(0.000319)	(0.001842)
Mental Health (eval 2000)	*10*	-0.0179	-0.0143	-0.025	-0.0133	-0.0151	-0.0056	-0.0341
	20	-0.0105	-0.0106	-0.0165	-0.012	-0.0111	-0.0056	-0.0333
	30	-0.0082	-0.0089	-0.0136	-0.0100	-0.0093	-0.0056	-0.0319
	40	-0.0076	-0.0077	-0.0103	-0.0085	-0.008	-0.0055	-0.0301
Median:	***50***	**-0.0064**	**-0.0065**	**-0.0093**	**-0.0071**	**-0.0068**	**-0.0055**	**-0.0278**
	60	-0.0051	-0.0056	-0.0083	-0.0059	-0.0058	-0.0054	-0.0243
	70	-0.0042	-0.0046	-0.006	-0.0048	-0.0048	-0.0053	-0.0207
	80	-0.0034	-0.0038	-0.0037	-0.0041	-0.004	-0.0052	-0.0171
	90	-0.0009	-0.0031	-0.0021	-0.0034	-0.0032	-0.0049	-0.0131
Mean:		-0.0075	-0.0075	-0.0108	-0.0045	-0.0082	-0.0053	-0.0240
(SE):		(0.000524)	(0.002281)	(0.000828)	(0.003951)	(0.001549)	(0.001751)	(0.009706)

Table 2.5: Marginal Effects. The Gumbel and the semiparametric binary estimates match closely, and for cigarettes, the semiparametric count, the structural model, and the negative binomial 1 (with constant dispersion) match as well. The last two count models, the negative binomial with mean dispersion (NB2) and the Poisson, fail to match the marginal effects of any of the other distributions, supporting the model. Standard errors are computed using the bootstrap technique with 50 replications.

Gumbel Model
Dependent Variable: Respondent reports stealing item > $50 at any point between 1997 and 2003

	Cig Usage	Demographic Controls					All	Main
	Only	Only	and Reg.	and Mental	and Work	and Wage		
Days smoking in	-0.040***						-0.034***	-0.035***
last month (age 16/17)	(-14.006)						(-10.561)	(-10.914)
Age in 1997		0.001	0.007	0.020	-0.001	-0.015	0.028	0.016
		(0.046)	(0.242)	(0.717)	(-0.024)	(-0.471)	(0.980)	(0.552)
Female		0.817***	0.818***	0.852***	0.818***	0.888***	0.822***	0.841***
		(9.590)	(9.598)	(9.806)	(9.600)	(8.840)	(9.392)	(9.670)
Black		-0.020	-0.027	0.195	-0.006	-0.021	-0.124	-0.039
		(-0.200)	(-0.254)	(1.818)	(-0.060)	(-0.176)	(-1.054)	(-0.383)
Hispanic		-0.142	-0.106	-0.015	-0.130	-0.114	-0.167	
		(-1.331)	(-0.933)	(-0.133)	(-1.208)	(-0.908)	(-1.432)	
Household Income		0.004**	0.004**	0.001	0.004**	0.003**	0.000	
($1000s)		(3.006)	(3.163)	(0.848)	(2.990)	(2.585)	(0.392)	
Northeast			-0.163				0.008	
(1997)			(-1.079)				(0.059)	
North Central			-0.201				-0.058	
(1997)			(-1.425)				(-0.472)	
South			-0.027				0.142	
(1997)			(-0.197)				(1.199)	
Urban Residence			-0.224*				-0.214*	
(1997)			(-2.358)				(-2.239)	
Reg. unemp rate			-0.194					
			(-1.269)					
ASVAB Math/Verbal				0.007***			0.005**	0.008***
(1000 pts)				(4.246)			(2.890)	(5.030)
Mental Health				0.085***			0.060***	0.067***
(eval 2000)				(5.635)			(3.943)	(4.500)
Education				0.171***			0.143**	
(age 16/17)				(3.924)			(3.288)	
Weeks worked/year					0.001	0.004*	0.000	
(age 16/17)					(0.738)	(2.216)	(0.144)	
Wage (first reported)						0.087		
(age 16/17)						(0.854)		
Days consumed alcohol in							-0.022**	-0.020**
last month (age 16/17)							(-3.239)	(-2.936)
Constant	2.275***	1.453***	1.628***	-2.133**	1.446***	0.935	-0.889	0.361
	(43.793)	(3.506)	(3.786)	(-3.290)	(3.490)	(1.082)	(-1.327)	(0.734)
N	4782	4782	4782	4782	4782	3460	4782	4782
Log Likelihood	-1797.239	-1827.671	-1822.335	-1782.931	-1827.395	-1346.914	-1703.819	-1716.034
Likelihood Ratio Statistic	173.3	113.1	124.4	203.5	113.0	103.5	361.9	335.3

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
t-statistics in parentheses

Table 2.6: A series of Gumbel models comparing covariates that might predict theft behavior. In keeping with the model, virtually nothing that related purely to variation in benefits (locational or regional dummies) has significant effect. Excluding gender, the most significant measures across specifications are those that speak to patience, discipline and forward-thinking.

Gumbel Model
Dependent Variable: Respondent reports stealing item > $50 at any point between 1997 and 2003

	All Respondents (with all variables)	Males Only	White Males Only
Days smoking in	-0.034***	-0.038***	-0.038***
last month (age 16/17)	(-10.563)	(-9.959)	(-7.458)
Days consumed alcohol in	-0.023**	-0.013	-0.021
last month (age 16/17)	(-3.251)	(-1.468)	(-1.769)
Age in 1997	0.037	-0.005	-0.076
	(1.266)	(-0.136)	(-1.593)
Female	0.823***		
	(9.405)		
Black	-0.128	-0.098	
	(-1.080)	(-0.680)	
Hispanic	-0.152	-0.251	
	(-1.298)	(-1.777)	
Household Income	0.000	0.000	0.000
($1000s)	(0.357)	(0.311)	(0.101)
Northeast	-0.119	-0.224	-0.407
(1997)	(-0.785)	(-1.217)	(-1.480)
North Central	-0.182	-0.199	-0.357
(1997)	(-1.286)	(-1.156)	(-1.405)
South	0.026	-0.043	-0.352
(1997)	(0.192)	(-0.262)	(-1.367)
Urban residence	-0.213*	-0.140	-0.075
(1997)	(-2.233)	(-1.235)	(-0.527)
Reg. unemp rate	-0.283	-0.300	-0.571
(age 16/17)	(-1.846)	(-1.620)	(-1.889)
ASVAB Math/Verbal	0.005**	0.006**	0.007**
(1000 pts)	(2.871)	(2.778)	(2.667)
Mental Health	0.060***	0.051**	0.044
(eval 2000)	(3.966)	(2.796)	(1.720)
Education	0.149***	0.132*	0.206**
(age 16/17)	(3.419)	(2.535)	(2.729)
Weeks worked/year	0.000	-0.001	0.001
(age 16/17)	(0.126)	(-0.804)	(0.236)
Constant	-0.958	-0.010	0.474
	(-1.424)	(-0.012)	(0.406)
N	4782	2416	1410
Log Likelihood	-1702.145	-1047.933	-570.053
Likelihood Ratio Statistic	361.5	194.7	133.8

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
t-statistics in parentheses

Table 2.7: Applying the final gumbel model of 2.6 to all respondents (where all variables are populated), male respondents only, and white male respondents only. As with 2.6, variables that seem most to likely measure costs of crime, particularly perception of future costs, show significant explanatory power. This supports the theoretical model and both structural models.

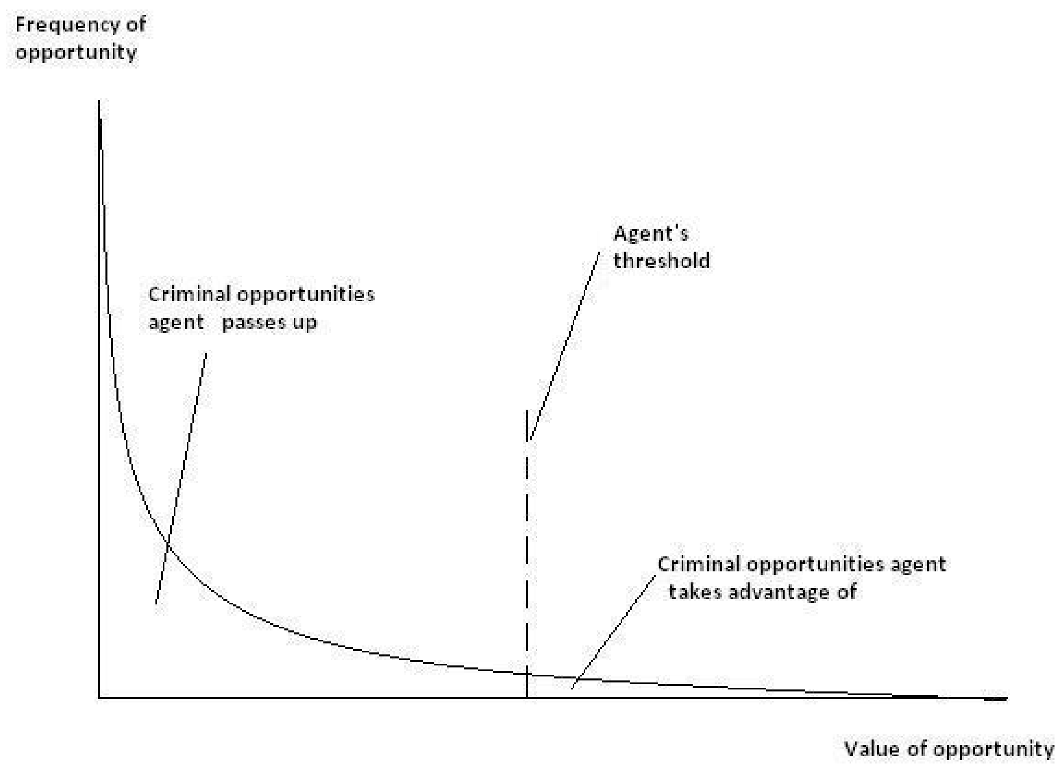

Figure 2.1: A simple visual explanation of the threshold model of this paper. In every time period (i.e., roughly daily) every individual draws a theft opportunity from a common lognormal distribution. If the opportunity offers a greater expected return than the expected cost of committing theft, the individual takes it (steals); if not, the individual lets it pass by.

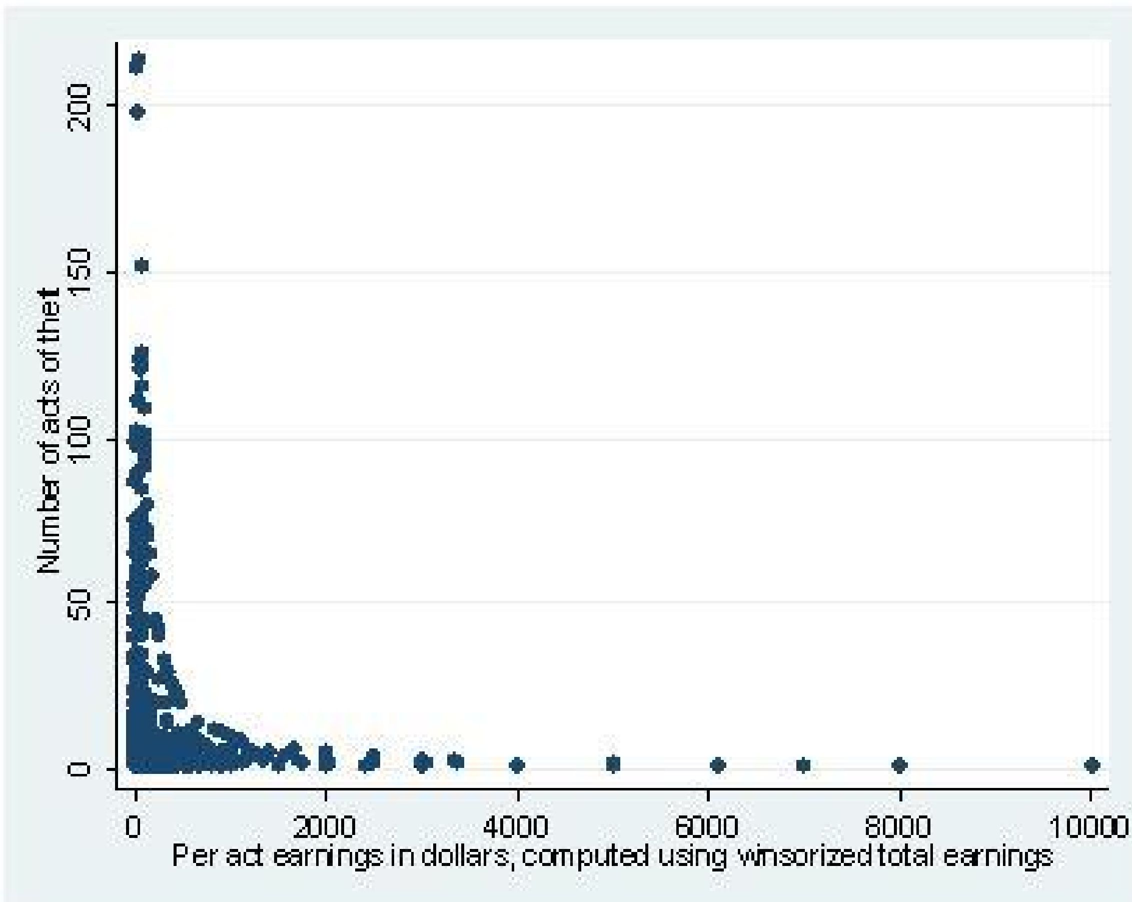

Figure 2.2: Data on earnings from theft and number of acts, after winsorizing average earnings per theft at the 95% level.

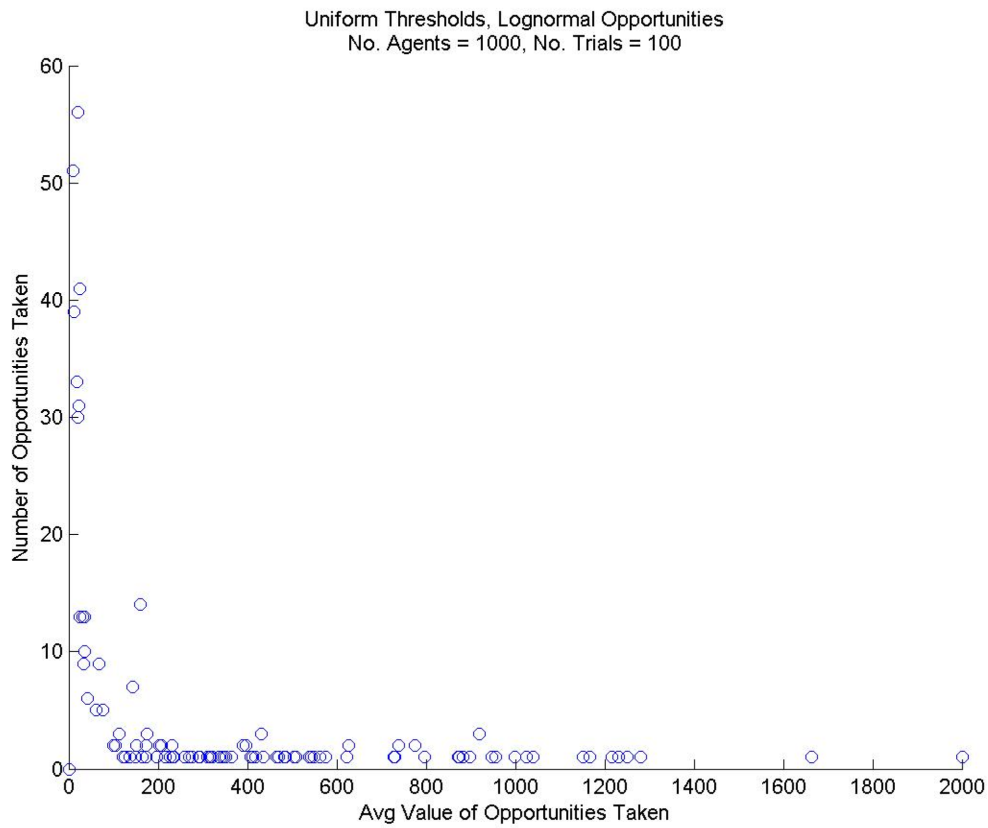

Figure 2.3: Outcome of a simulation of the model, with thresholds uniformly distributed from 0 to 2,000, and opportunities lognormally distributed with 0 mean and variance 1. Note that the simulation data shows a very strong similarity to the data graphed in Figure **??**, suggesting that theft opportunities follow a lognormal distribution. When the distribution of opportunities is changed (to, for example a uniform distribution) this similarity is not apparent.

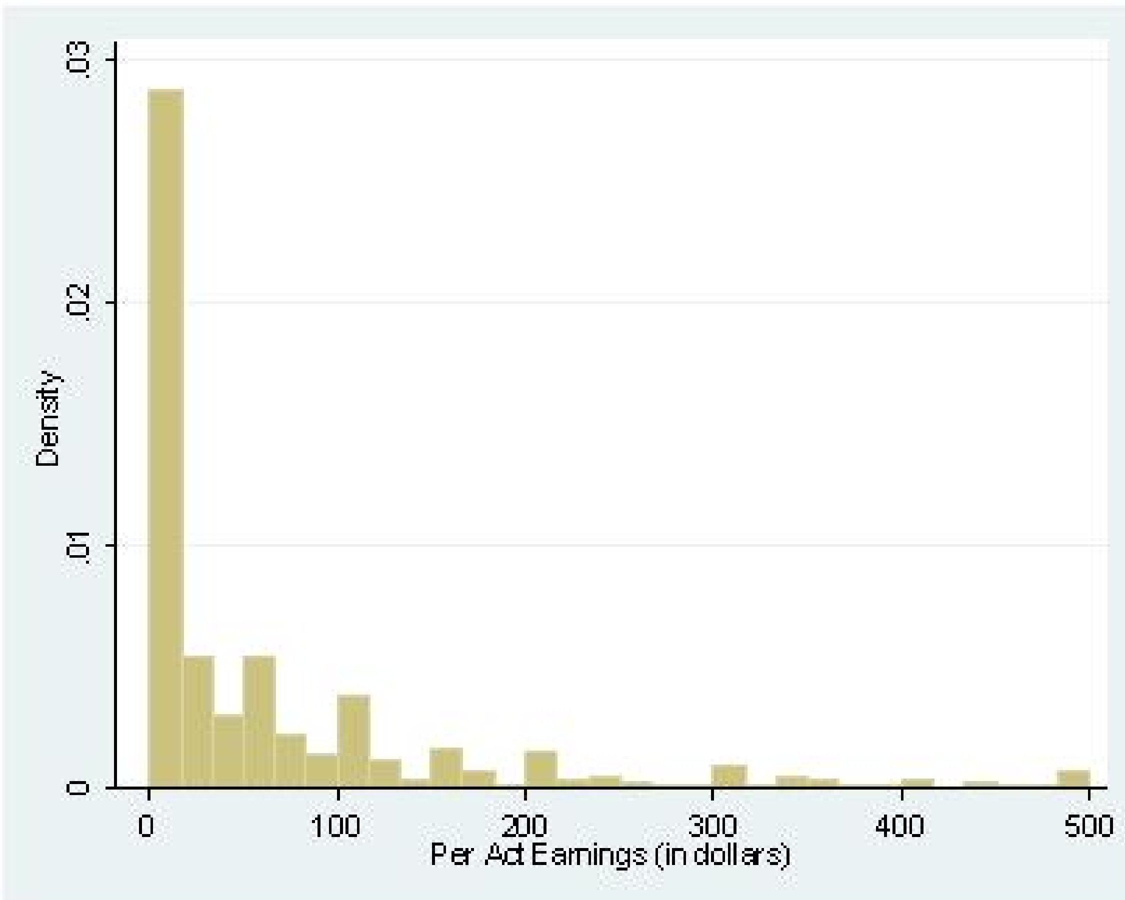

Figure 2.4: A histogram of winsorized per-act earnings for respondents who report theft, excluding thieves who report per-act earnings greater than $500. Ninety percent of all respondents who report theft are included in this histogram.

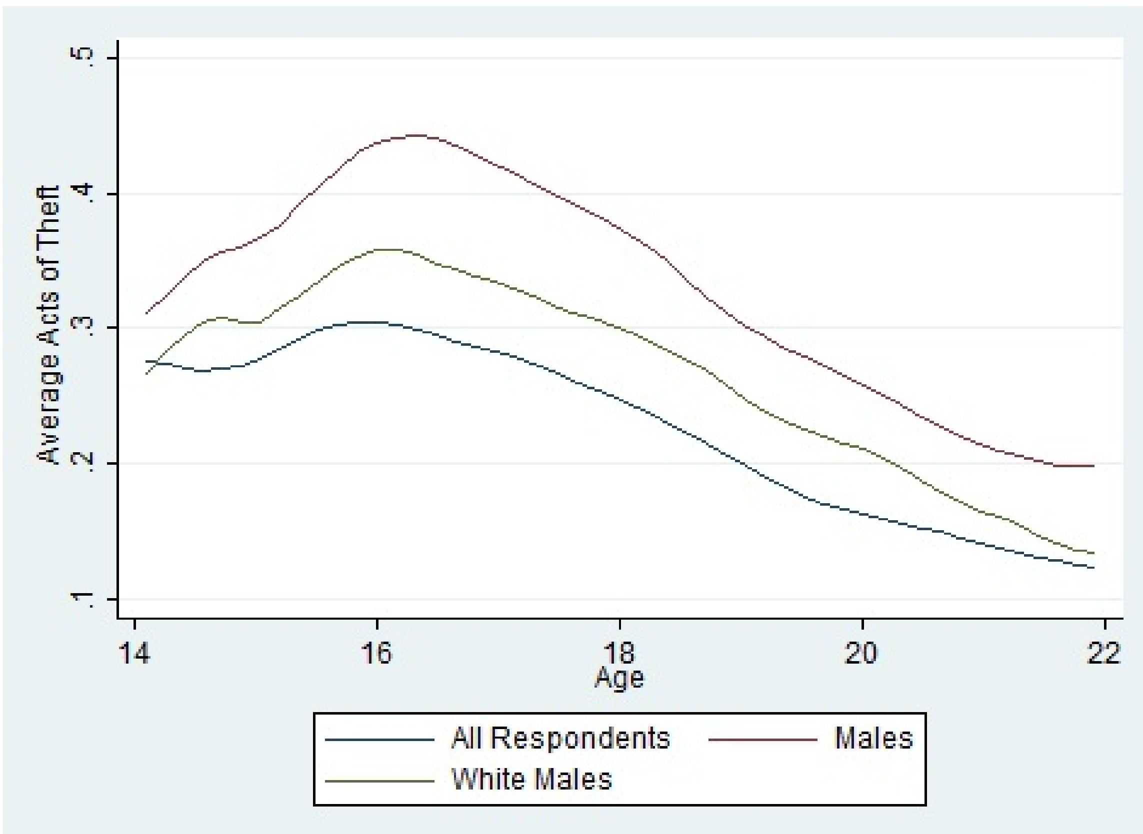

Figure 2.5: The aggregate age pattern of theft in the NLSY 1997 data for years 1998 to 2003.

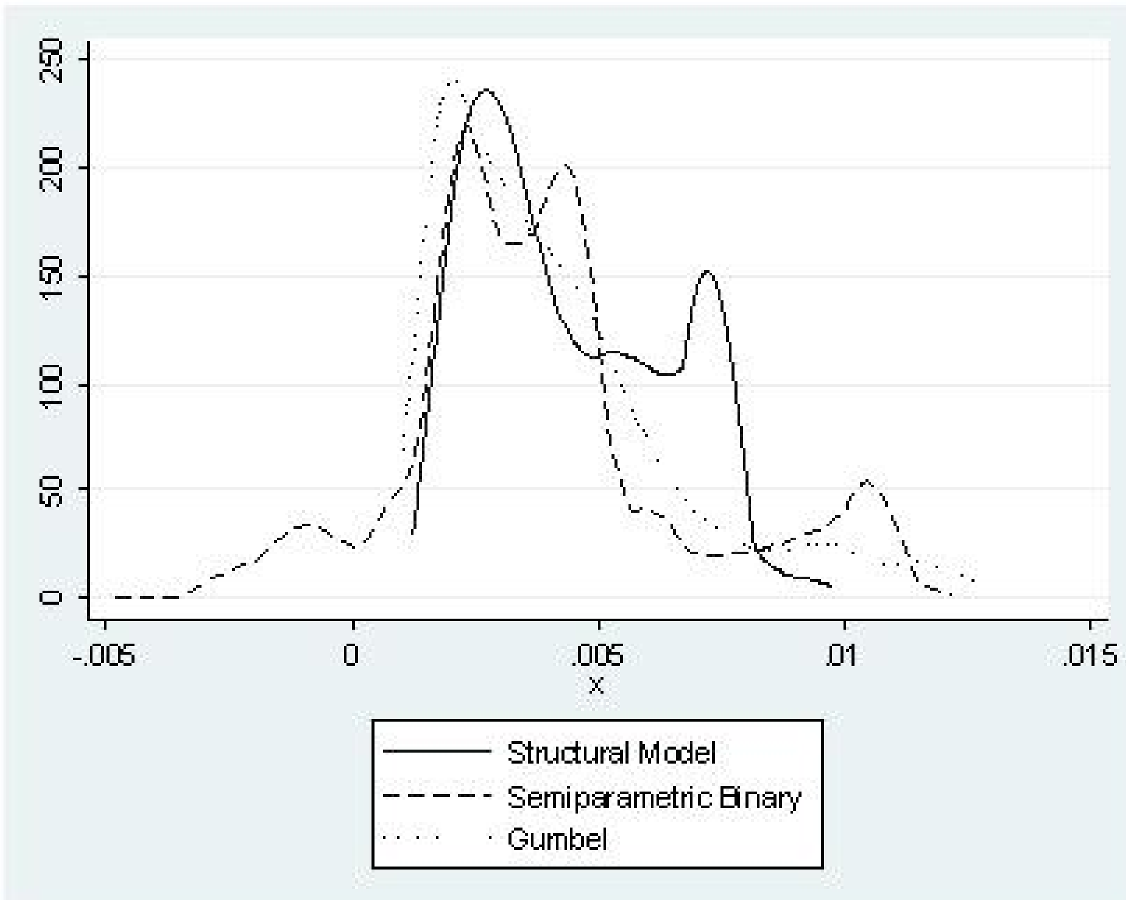

Figure 2.6: The distribution of marginal effects for smoking behavior for the Gumbel distribution (binary discrete), the semiparametric binary discrete estimation and the structural model (bivariate, with counts and total theft earnings). The structural model incorporates other information and so the marginal distribution is slightly different, but is in agreement.

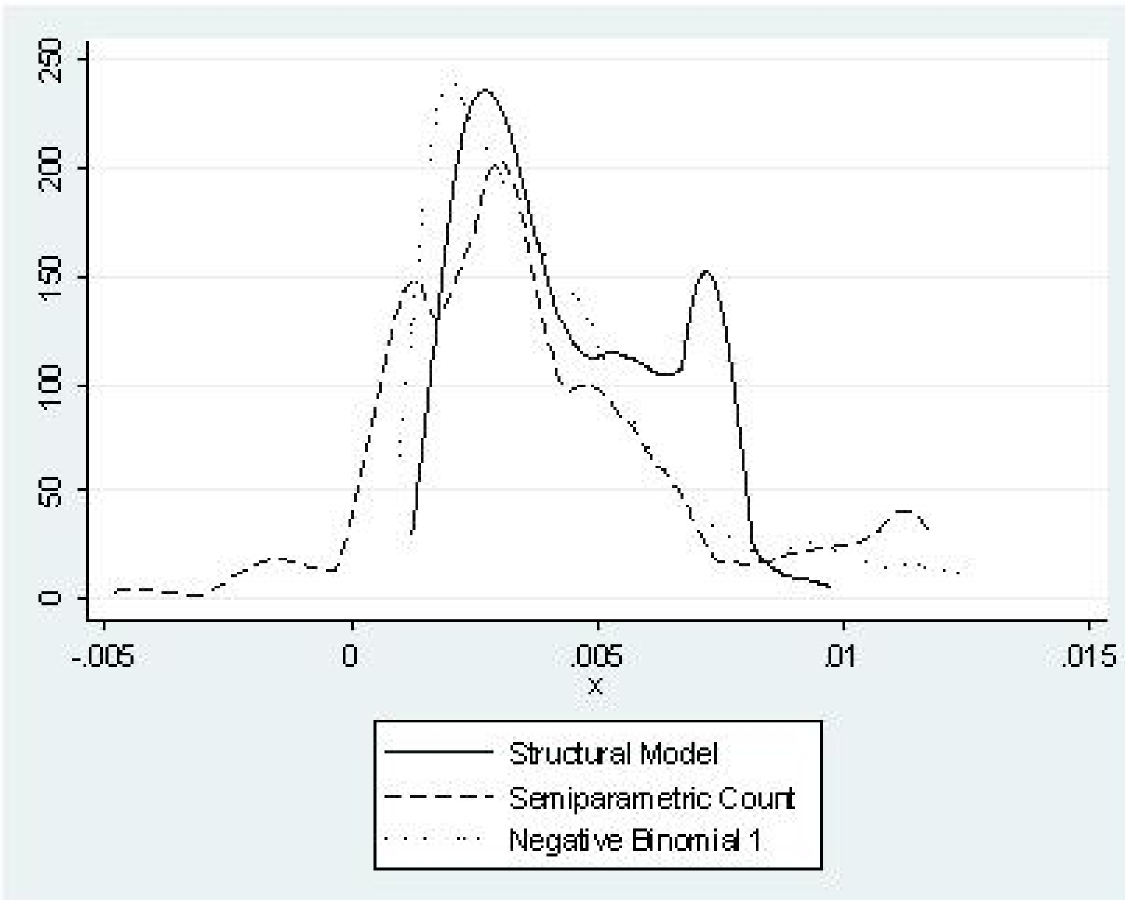

Figure 2.7: The distribution of marginal effects for smoking behavior for the negative binomial distribution with linear (constant) dispersion (count with unobserved heterogeneity), the semiparametric count estimation and the structural model (bivariate, with counts and total theft earnings). The structural model incorporates other information and so the marginal distribution is slightly different, but is in agreement.

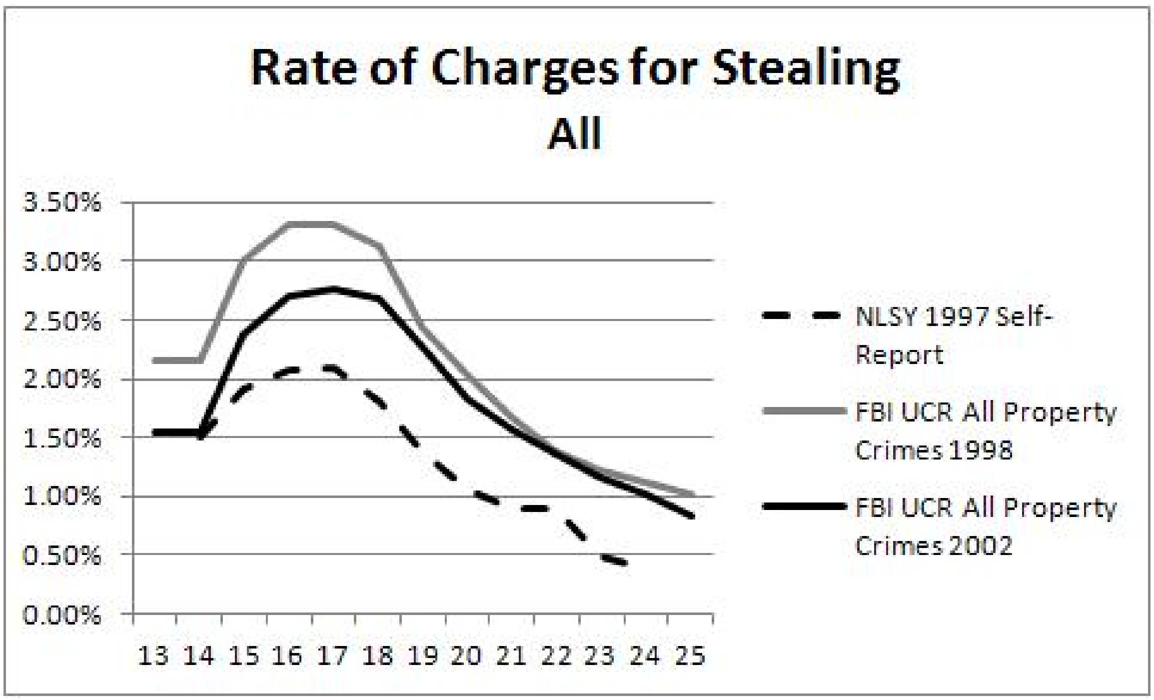

Figure 2.8: Self-reported rates of arrest among NLSY respondents compared with FBI Uniform Crime Reports of property crime for 1998 and 2002. Self-reported arrests seem to be a reasonably close fit with official records.

Chapter 3

Criminal Careers in Two Dimensions: Age-Specific Inclination and Individual-Specific Intensity

A good run on the street was six months, and you had to have a clear head and a lot of self-confidence to make it even that long.

Clockers, Richard Price

3.1 Introduction

The noun "criminal" has two standard definitions. In the Oxford English Dictionary they are given as "2. (a) A person guilty or convicted of a crime. (b) A person with a tendency to commit crime."[1] While there is a natural tendency to lump the two ideas together, here is a world of difference between them.

An illustration of this difference and the complexities we are faced with as a result can be seen in the publicly available story of Pedro Hernandez as of June 2012. In May 1979, six-year-old Etan Patz disappeared on his way to school in New York City. The case became national news, and led to radical changes in awareness of missing children, and expanded efforts to protect children from abduction. Thirty-three years later, in late May 2012, Pedro Hernandez confessed to having murdered Patz on the day of his disappearance, and Hernandez was charged with second-degree murder. At the time of the alleged murder Hernandez was 18 years old and worked at a bodega near Patz's home in the Manhattan neighborhood of SoHo. While as of this writing the police investigation is ongoing and there is some

[1] Other dictionaries give very similar definitions.

potential that Hernandez's confession may be false, brought on by mental illness, it seems likely that Hernandez committed the murder[2]. If so, and if (as also seems likely at the moment) he is convicted of a crime, then Pedro Hernandez is without a doubt a criminal in the first sense of the word; obviously the murder of a six-year-old in cold blood is one of the most extreme crimes imaginable.

At the same time, in the news coverage and in the publicly available facts on Mr. Hernandez, there is almost no evidence that he is a criminal in the second sense of the word. Before May 2012, when he was charged with the murder of Etan Patz, he does not seem to have had any criminal record at all. This is not to suggest he was a model citizen. He has had serious mental problems and psychotic symptoms for most of his life. He is in his second marriage and both his ex-wife and his present wife took out protection orders against him. No accusations of actual domestic violence have yet surfaced with regard to either marriage, and it is possible that the protection orders were temporary, based solely on a fear of potential violence.[3] There is certainly evidence that Hernandez could be aggressive and intimidating: His second wife was charged with fraud, and the victim of the fraud, who had persuaded the police to press charges against the wife, says Hernandez called her up and threatened her. All of this suggests issues of anger and self-control, and it is possible that worse actions will come to light. But given Hernandez' well-documented mental problems and his confession of murder there is surprisingly little evidence that he possessed significant ongoing criminal *tendencies*.

Hernandez' story is thankfully unique in many ways, but the general pattern of a brief period of criminal activity - a spike - with no activity before or after is very common in data on criminal behavior. There is much more inactivity than activity across the vast majority of criminal careers, and criminal activity is usually limited to a few periods or

[2] The discussion of Pedro Hernandez is based on reports in the New York Times from May and June 2012, particularly the article "For Patz Suspect, a Tumultuous Past", June 10, 2012.

[3] Temporary restraining orders can be issued by a judge who believes there is *risk* of abuse or violence, but final restraining orders that can only be issued with evidence that violence has actually occurred. For both marriages, Hernandez seems to have been a resident of New Jersey, so I am describing New Jersey law.

less. These fluctuations have been discussed by earlier researchers; Nagin and Land (1993, p. 334) observe that “discontinuous jumps from a state of zero criminal potential to positive criminal potential” represent “the key point of controversy” in understanding criminal careers.

The argument of this paper is that when aggregating across types of crime, criminal careers vary systematically along only a single measure, which might be termed intensity (or frequency).[4] This univariate difference between offenders has been largely obscured because of the spiky, intermittent nature of criminal activity. For individuals with the same intensity levels, differences in the timing of criminal behavior are likely to be due to contingent or perhaps even random influences. Other measures, such as group-based trajectory models, that try to discriminate further in criminal careers beyond intensity levels are not actually getting to meaningful, stable variation between individuals.

One complicating factor is that for individuals committing multiple crimes, there is some temporal link between crimes; they are not stochastically independent. That is to say, individuals’ criminal activity occurs in “runs”. Even knowing the individual’s age and level of intensity, our odds of successfully predicting whether or not an individual will commit a crime in period t are improved if we know whether or not the individual committed a crime in period $t-1$, where periods are roughly 3 months to 2 years. This could either be due to autocorrelation over time, or alternating periods of activity and inactivity. This particular issue requires further research[5].

In effect I am revisiting some of the analysis in Greenberg (1991, see particularly the section

[4]This article is focused on analyzing general propensity to commit crime, as opposed to understanding specific choice of crime A vs crime B. It is quite likely that there may be issues involved in specialization, such as inclination to violence or pyromania, which would help us to predict variation in *type* of crime. Such analysis is beyond the scope of this paper.

[5]This point of the temporal relationship between acts of crime is somewhat similar to the idea of intermittency developed by Nagin and Land (1993). As implemented by Nagin and Land, intermittency includes measures of the age-crime curve and individual-specific intensity, while I believe it is both possible and advisable to keep all three ideas separate. See the discussion of the model in section 3.3.

"The Age-Crime Relationship", pp. 33-36). The probability of an individual committing a crime in a particular period seems to be defined by three things: first, the general probability of any human individual committing a crime at a particular age (summarized in the well-known "age-curve"), second, the specific level of criminal intensity of the individual in question, and third, the complicating factor of the temporal link between crimes.

The paper is structured as follows. In section 3.2 I review the literature on criminal careers. In section 3.3 I present a mathematical model that tries to link some of the different viewpoints on criminal careers. In section 3.4 I then review some of the patterns in offending that are visible across three different data sets. Importantly, all three of these data sets include non-offenders, offenders with only one or two recorded acts, and offenders with a large number of acts, thus minimizing sample selection issues. I find that a very large percentage of criminals are active for only a few (sometimes just a single) short periods, and that for all but a very tiny minority there are more periods of inactivity than activity. In section 3.5 I briefly revisit how different researchers have looked at this spikiness or intermittency, and how well their analysis fits the data. In section 3.6 I compare the distribution of offenses over the early life-course (early teen years to mid-to-late twenties) between offenders with high activity levels vs offenders with low activity levels. Controlling for total activity, the distribution of the activity - roughly the trajectory of the activity - seems almost identical across the groups. The only difference in relative distribution between high activity and low activity individuals is that high activity individuals show slightly more relative activity in later periods. While the details change, the broad pattern holds across self-reported behavior, arrest records, and conviction histories. I do not look at incapacitation effects directly, but it does not appear that incapacitation radically changes the picture. In section 3.7 I discuss the implications and areas for future research.

3.2 Background

There have been a number of excellent reviews of the history of criminal taxonomies and the criminal career paradigm in the last decade. The reader is directed to Sampson and

Laub (2005), Nagin and Tremblay (2005) and Brame et al. (2011) to see some recent very thorough discussions. An excellent summary of work in the 1980s and 90s is Piquero et al. (2003). A brief summary of recent (the last 40 years) work follows, focusing on how the fluctuations in criminal behavior are handled by different researchers and incorporated into their models of criminal behavior.

3.2.1 Focus on Desistance and Recidivism

A number of researchers have viewed the fluctuations in criminal activity as one of uneven desistance or recidivism. While not expressly stated, there seems to be an assumption that criminal or delinquent careers all begin at roughly the same age but vary greatly in when they end. Thus both the age-crime curve and the spikes in criminal behavior are implicitly explained as a large "pool" of juvenile delinquents becoming active in their early teens that then slowly and unevenly "dries up" over the late teens and early twenties, leaving only the most dedicated criminals by the late twenties.

For studies of delinquents, the logic of this is straightforward; as pointed out in Tracy et al. (1990, p. 176), the age-at-onset of delinquency is bounded at the top, at age 17-18, by statute. Thus virtually all delinquents have an age-at-onset between age 7 and age 17, and the vast majority within a few years in the early teen years. Since it is assumed that a substantial number of delinquents become adult offenders, the transition to "non-offender" status can take place at any age from seven onwards. This explains much of the analysis of Tracy et al. (1990), with its focus on the probability of offense $n+1$, conditioned on the commission of the first n offenses.

For criminal behavior generally, looking at both juvenile and adult offenses, this limit does not hold. In the three datasets I have looked at, offenders who commit only one act vary greatly in when they commit that single act - from ages 10 to 30. To be clear, it is not that any researchers have explicitly denied the possibility of activity occurring only later in life. However, there are no models I have come across that specifically discuss or include such patterns, even though they seem quite common in the data.

For the statistical model expounded by this paper, social control theory has the great advantage of focusing on why individual offense rates might vary from period to period. Some of the classic analyses have accepted the model of the delinquent who then desists (see Sampson and Laub (1990), Sampson and Laub (2003)), but other papers in this school such as Sampson et al. (1997) have looked at cross-sectional differences that explain why a potential offender at any age may choose to act or not.

3.2.2 Adolescent-Limited and Life-Course Persistent

In a seminal article, Moffitt (1993) proposed a taxonomy that divides criminals into the broad categories of "life-course persistent" and"adolescent-limited" criminals who are only active for a short period. This taxonomy explains the intermittent behavior as a composition of two groups. The "adolescent-limited" offenders engage in criminal behavior briefly, usually in their teen years, and then stop. The "life-course persistent" offenders have a roughly flat career path, but because they do it for temperamental reasons rather than for any goal-directed purpose, there is some natural fluctuation in their behavior.

3.2.3 Criminal Propensity

In their *General Theory of Crime* (Gottfredson and Hirschi, 1990) and other work, Gottfredson and Hirschi emphasize the role on individual-specific characteristics that are stable over time - particularly lack of self-control and impulsivity. In their early discussion of the criminal career paradigm (Gottfredson and Hirschi, 1986) they seem to assume that a "career" requires planning and stability; in this view, the fluctuations in criminal activity are a natural byproduct of the disorganization and lack of planning common to individuals with a propensity to commit crime. In their discussion of the age-crime curve (Hirschi and Gottfredson, 1983, see p. 582) they similarly view the effects of age as "noncontroversial and theoretically uninteresting"; it appears that they mean that individuals offend more in adolescence than in childhood or adulthood.

The analysis of Greenberg (1991) shows that in fact the Gottfredson and Hirschi view matches aggregate cross-section data well. Towards the end of this article Greenberg suggests and does preliminary analysis for a model that is extremely close to the model of this paper - a lambda that is combination of an individual-specific component (as with Gottfredson and Hirschi's ideas) and an age-variant component (as is used in much of the criminal career analysis).

3.2.4 Group-Based Trajectories

Into the debates on criminal careers and criminal propensity discussed above Nagin and Land (1993) introduce the group-based trajectory methods (GBT, GBTM, or TRAJ) that have since dominated the discussion of patterns of criminal behavior over the life-course. The major advance of this work is an explicit, formal statistical technique that allows researchers to analyze and summarize a wide range of behavior patterns across an extensive range of data. The estimated trajectories are generally very smooth and graceful - see Broidy et al. (2003), for example. The technique has been used a great deal in criminology but also other disciplines; a summary of its success can be found in Nagin and Odgers (2010). The technique can be found in a number of general econometric textbooks; see Greene (2008) and Cameron and Trivedi (1998).

The approach of group-based trajectories has become a standard approach to understanding criminal careers[6] and is indeed considered a model latent-class statistical technique, cited in numerous textbooks (see Greene (2008) and Cameron and Trivedi (1998)). At the same time, there is still significant controversy. For example, Skardhamar (2010) and Sampson and Laub (2003) (also Skardhamar (2009), Sampson and Laub (2005)) have recently critiqued the approach for constructing discrete categories on what may be only continuous variation.

[6]While there are clearly different perspectives on this issue in Brame et al. (2011), Nagin and Tremblay (2005), Sampson and Laub (2003) and Sampson and Laub (2005), the dominant role of the method is undisputed.

3.3 Statistical model

A criminal career is generally measured with two variables. The first is λ_{it}, the criminal activity perpetrated by individual i in period t, and the total crime perpetrated by the individual across all periods of observation,

$$\lambda_i = \sum_T \lambda_{it}$$

.

Combining the observations of the different schools of thought summarized in the previous section, it seems likely that there are three components that define the levels of λ_{it} and λ_i.

The core component is stable individual-specific traits (what Wilson and Herrnstein (1998) would call "constitutional factors"), which we can summarize as

$$p(\theta_i, X_i)$$

where θ_i is unobservable traits and X_i is observables. Perhaps the best way to explain this is as the default probability of individual i committing a crime in any particular instant of his life (or in any instant in the interval that we observe him), which is then scaled upwards and downwards by the two other components, discussed below.

The second component is the age-crime curve effect that is generally acknowledged by all researchers (as Hirschi and Gottfredson (1983) write, a "noncontroversial" pattern) which we can summarize as

$$g(a_i(t))$$

This can be viewed as a factor that scales up $p()$ as the individual moves chronologically from birth to age 17, and then scales it down over time. The exact value of $g(a_i(t))$ for any

t we leave open for the moment; it appears to vary across data sets and across crimes.

Finally there are environmental factors, which we can denote with

$$\varepsilon(Y_i(t), \omega_i(t))$$

where $Y_i()$ are time varying observable variables and $\omega_i()$ are time varying unobservable variables. The function fluctuates stochastically over time, in a way that is unpredictable in period 0. As discussed in the introduction, the evidence strongly suggests that these variables are autocorrelated from moment to moment.

Working in continuous time, for any period s of duration δ

$$E(\lambda_{it}) = \int_s^{s+\delta} p(\theta_i, X_i) g(a_i(t)) \varepsilon(Y_i(t), \omega_i(t)) dt$$

and

$$E(\lambda_i) = \int_0^T p\left(\theta_i, X_i\right) g(a_i(t)) \varepsilon\left(Y_i(t), \omega_i(t)\right) dt$$

Understanding $\varepsilon()$ requires substantial, period-by-period data on environmental influences; nothing in the three data sets seems to offer serious insight into this. The most powerful of the three is the NLSY 1997 data set, but there do not seem to be many time-variant variables that show explanatory power, and those that do (smoking behavior for example) are strongly endogenous. Thus, for the purposes of this analysis, we are limited to only some basic observations on $\varepsilon()$ - which is not to imply that it is not an extremely important area of investigation.

It is worth considering how some of the standard views of criminal behavior can be matched to this model.

The group-based trajectory models can be understand as focusing on $p(\theta_i, X_i)\, g(a_i(t))$, and assuming some level of covariance between the two functions. That is to say, depending on certain unobservable traits θ_i, an individual may have a particular trajectory against the overall age-crime curve $g(a_i(t))$. The idea of intermittency can be thought of as adding an autocorrelation component to $\varepsilon\,(Y_i(t), \omega_i(t))$, albeit one that is not constant and that varies according both age and constant individual characteristics. I discuss intermittency further in section 3.5.3.

Social control theory might be understood as focusing on measures of $\varepsilon\,(Y_i(t), \omega_i(t))$. The focus on desistance as something explained by relationships could be interpreted as minimizing the influence of $g(a_i(t))$, relative to other theories. That is to say, social control theory seems to look at specific explanations for the drop in crime from 17 onwards, while other theories, at least implicitly, assume much of this is a "natural process.".

In contrast, the theory of Gottfredson and Hirschi could be interpreted as assuming that $\varepsilon()$ has a limited role, and that $p(\theta_i, X_i)\, g(a_i(t))$ explain criminal careers completely. We could even normalize

$$\int g(a_i(t))\varepsilon\,(Y_i(t), \omega_i(t))\, dt$$

to one, and set $p(\theta_i, X_i) = \lambda_i$.

In the remainder of the paper I will review specific data to see if we can understand a bit better how these three components fit together.

3.4 Basic career patterns

The purpose of this section is to review some of the general patterns in criminal careers that appear consistently across a number of datasets. The most important pattern for the thesis of this paper is the spikiness of criminal behavior: not only is the modal criminal career very short (one or two years) but the median crime is committed by a criminal with a career lasting 6-7 years but with activity in only 3-4 years.

I look at data from the Farrington and West data on London delinquents, the Wolfgang et al. study of a Philadelphia cohort, and the NLSY 1997 cohort. Table 3.1 gives an overview of the three data sets. In a series of tables below, I break down the percentage of individuals who are active criminals by number of years they are active and by the length of career.

While the datasets I am working with cover an extensive period of time (the shortest is the NLSY data, which goes 12 years from 1997 to 2009) none of the datasets track offenses after age 30. Thus we are artificially losing offenses for some individuals with criminal activity beyond their late 20s. However, since few of the individuals in any dataset show careers that come close to the full range of the data or close to the ending cutoff, it is unlikely that this effect radically changes the following analysis.

I begin by looking at the conviction data of Farrington and West. Table 3.2 summarizes the number of offenders who are active (i.e., are convicted) in a single year, who are active in two years, who are active in three years, etc. Several things are worth noting. First, as has been noted by earlier researchers, a large number of offenders are active for very brief periods - one-third of offenders for one year, the majority of offenders in only two years. Second, the median offense is committed by an individual who is active for four years (out of 15 years tracked in the data).

In table 3.3 I break down the Farrington and West data by the career length of the offender. It can be seen that criminal activity is dispersed over a range of years; the median criminal shows activity over 3 to 4 years, and the median offense is committed by an individual who shows activity over 7 to 8 years. This dispersion of activity across a length of time, as opposed to a contiguous block, appears in every data set.

Offenders frequently escape conviction for their offenses, so it is possible that the spikiness is simply an artifact of the reality that conviction is a high measurement threshold.

To compensate for this I now look at arrests in the Wolfgang dataset on Philadelphia delinquency. First, in table 3.4 I break down the Wolfgang offenders by number of different years in which they were arrested. The general pattern is the same as table 3.2. Nearly half of the offenders are only active (in this case, arrested) in a single year. Two thirds of the offenders are arrested in only 2 fewer years, and the majority of the offenses are committed by offenders who are active in 4 years or fewer.

Sorting the Wolfgang offenders by career length in table 3.5 we can see that while the median criminal shows activity over a 2 year span, the median crime is committed by a criminal active for nearly 7 years.

There is thus a very strong match between the pattern of criminal behavior spikes in the Farrington and West data and the Wolfgang data.

Since not all offenses trigger arrests, I next look at self-report data, to see if the intermittency continues. Using the NLSY 1997 cohort panel, with the data from 1997 to 2009, I look at self-report drug-dealing and theft of objects worth more than $50.

In table 3.6 I sort the NLSY data on self-reported drug-dealing by number of years active, and in table 3.7 I sort by length of career. The patterns vary only slightly from that seen in the previous two data sets; half of all offenders are only active in a single year, the median drug deal is committed by an offender reporting activity in just 4 years, but that activity is dispersed across roughly 6 years.

Looking at theft in the NLSY 1997, in particular focusing on self-reported theft of items worth more than $50, we see the most unique pattern, and by far the most spiky. In table 3.8 I sort NLSY theft offenders by number of years active. The substantial majority of offenders report activity in only one year, and virtually all of them (95%) report activity in three years or fewer. Breaking down NLSY theft offenders by career length in table 3.9 the pattern changes only slightly - nearly 95% of all thieves show activity in 6 or fewer years.

In summary, there is a great deal of spikiness or intermittency in criminal behavior. The median criminal shows activity in only one or two years, but the median offense is committed by a criminal active across 6 or 7 years, but only active in 3 or 4 years. For all but the most active criminals, the criminal career CV has a great deal of white space. If your modal criminal were to apply for a job with a criminal organization, he would have a lot of explaining to do. Any theory of criminal behavior (whether a taxonomy, model of propensity or criminal career) needs to directly engage this spikiness.

3.5 Responses to Criminal Behavior Spikes

In this section I review several responses to the pattern of spikes. A number of them are from the existing literature (some previously touched in in section 2).

3.5.1 Selection Bias or Data Noise

One of the most obvious responses to the fluctuations in activity is that there is an issue with selection bias or noise in the data.

In the conviction and arrest data there are detection issues, in that only a subset of crimes are observed by the police. Potentially criminals are active for more extended periods, but they are only occasionally caught, thus creating the illusion of spikes in behavior. This is certainly a feasible explanation, and it seems likely that a number of the individuals who are being tracked in the Farrington and West and Wolfgang datasets are indeed only being caught some of the time. However, there is a significant problem with it: the self-report data shows significantly *greater* spikiness. If detection was driving spikiness in the arrest or conviction data, we would expect it to drop in the self-report data but the reverse happens.

There is a similar but contrasting explanation for the spikes in the self-report data. For the self-report data, individuals with only mild criminal tendencies may be more honest or more trusting. Because they are not hard cases they are more willing to honestly discuss their behavior in interviews. Chronic recidivists, on the other hand, are by definition less honest, and may also distrust authority figures more. As a result, they don't report their activity to the NLSY. This explanation is also quite likely to be partially true and there is some evidence of underreporting, perhaps as much as 30%, given a small but substantial number of respondents who report being charged with crimes but deny committing them.

But the fact remains that no matter what technique we use to measure criminal behavior, we find a large numbers who are active only briefly, and even more who are only active part of their career. How do we handle these individuals as we try to understand criminal behavior and criminal careers?

3.5.2 Focus on High Activity Offenders

One practice that tends to limit the intermittency issue is to exclude low activity offenders, by focusing on individuals with significant records. This was one of the criticisms that Gottfredson and Hirschi (1986) level against the criminal career analysis of Blumstein and Cohen (1979).

While ignoring low activity offenders may be justified in certain situations where the researcher is focused on a particular question, there is no clear general justification for it. There is strong evidence that they are committing real crimes: individuals who report drug dealing and theft in only year of the NLSY are not clearly of less interest or relevance than individuals reporting several years of activity.

The point is even clearer when we look at serious crimes. The discussion of Pedro Hernandez's story at the beginning of this paper, while clearly anecodtal, gives some sense of how much damage a single offense can do. Moving to more comprehensive analysis, Lisak and Miller (2002) gathered self-report data on the commission of rape from anonymous men. A summary of some of their data is shown in table 3.10. As can be seen, while the majority of rapes recorded in the dataset were committed by a minority of rapists with multiple offenses, there are a substantial number of rapists who (to the point of the survey, at least) have only committed a small number of rapes. If these rapists were to desist after a small number of rapes (as seems likely for at least some cases), they would still be rapists, and our interest in them from a policy or analytic perspective would remain.

3.5.3 Intermittency Variable or ZIP

If there are such strong spikes in the Farrington and West data, why do these spikes not show up in the trajectories estimated in Nagin and Land (1993) or Nagin (2005)? The answer is that in the Group-Based Trajectory method, fluctuation in activity is generally dealt with in one of two ways: via an intermittency parameter, or via zero-inflated Poisson.

The intermittency parameter is π in the original paper Nagin and Land (1993), described on page 334. Citing the issue of intermittency in Barnett et al. (1989), the authors write "[t]he individual is assumed to be active in during all of t with probability π or, alternatively, to be inactive for all of t with probability $1 - \pi$. During periods of activity $P(N; \lambda_i)$ describes the probability distribution of the number of crimes committed by individual i. Thus, an individual may be active but still have no recorded offenses. During periods of inactivity, however, the probability of crime is strictly 0." In the estimation procedures, π is a linear function of four binary measures that predict criminal behavior (summarized as TOT), age and age squared (described page 344). The parameters of the linear function are virtually all strongly significant.

In later work (Nagin, 2005), the same dataset has been analyzed by Nagin using zero-inflated Poisson (ZIP) methods. In the ZIP estimation procedure, two processes are posited - the first is the Poisson process, the second is effectively a censoring process that only allows some observations to register a Poisson draw. As Nagin rightly notes, the ZIP ends up playing an "analytically equivalent" role to the intermittency procedure (page 34).

Although π is estimated as a conditional sample mean; since there is never an ex post examination of the range of actual values π, we never understand the potential role of outliers or even different ranges of π. Using the Farrington and West data and the parameters of the second stage analysis of Table 6 (Nagin and Land, 1993, p. 349) I generate values of π for all individual/period combinations up to age 24 (a total of 3,234 observations). Figure 3.1 shows the results. The distribution of values is fairly scattered, without a clear central mass; it is not clear that π is truly measuring an underlying population value or is simply a "sink" for unexplained variation.

3.6 Sorting by Intensity

This section looks at offenders in the three data sets, sorted by total activity (one year, two years, etc, or one act, two acts, etc). It first looks at the distribution of this activity across the life-course, and finds that high intensity offenders show a similar distribution of their activity to low intensity offenders. It then looks at population means of various attributes, and shows that intensity levels do a good job of predicting underlying differences in attributes, while the timing of activity (i.e., early adolescence vs adulthood) does not.

3.6.1 Changes in Relative Probability of Committing a Crime

How does the distribution of activity of an offender who only acts in one year compare with that of an offender who is active across many years? That is to say, we know that a high intensity offender will commit more crime (by definition) but is there a difference in where this crime comes? To understand the issue better, consider a few possibilities suggested by the literature.

One of the well-established patterns of criminology is that age of first offense is a good predictor of how long and how serious a criminal career will be. One possible explanation of this pattern is that low intensity offenders only *extremely rarely* commit a crime before age 15, while high activity offenders are just as likely to commit crime at age 13 as they are at age 18 as they are at age 23.

Alternatively, consider the taxonomy of Moffitt (1993) (see especially figure 3 on page 677), which can be interpreted as predicting activity centered around age 17, the peak year of delinquency. Perhaps low intensity offenders *only* commit crime between ages 16 and 18?

As a first step in exploring this question, in figure 3.2 I present kernel regressions of activity for the Wolfgang data, sorted by number of years active. The lower curve shows the probability of a low intensity criminal being arrested in that year, while the higher curve shows the probability of a higher intensity criminal (anybody with 2 or more years of activity) being active in that year. It is clear that the general pattern is similar, in that both curves increase and decrease across the same age ranges. However, because the curves show absolute levels of activity, it is difficult to get to the question of relative change in activity.

In order to compare the distribution of criminal behavior across different ages in low intensity offenders, mid intensity offenders, and high intensity offenders, I scale the activity of each individual i in year t by their total activity. I do this using two measures of activity, the first looking at years of activity, and the second at acts.

For years of activity, I use the measure a_{it}/A_i, where

$$a_{it} = \begin{cases} 1 & \text{if } i \text{ is active in year } t \\ 0 & \text{if } i \text{ is active in year } t \end{cases}$$

and

$$A_i = \text{total year of activity for individual } i$$

For example, the single year a low intensity offender commits a crime is coded with a 1.00, while each of the three years a mid-intensity offender commits a crime are coded with 0.33, and the five years a high intensity offender is active is coded with a 0.20. This allows us to abstract away from the higher rate and more numerous years of activity of the high intensity offender to see if there is a difference in where this activity occurs in the life course.

For total acts I use the measure λ_{it}/λ_i where

$$\lambda_{it} = \text{number of acts by individual } i \text{ in year } t$$

and

$$\lambda_i = \text{total number of acts by individal } i$$

For example, the single year a low intensity offender commits a single crime is coded with a 1.00, and all other years would be coded zero. For an offender who commits three crimes in two years, the year with two acts would be coded 0.66, and the year with one act would be coded 0.33, all other years zero.

Figure 3.3 shows this for the Farrington and West data. I group offenders into three rough categories - offenders who are active for one year, the particularly spiky offenders, offenders who are active for two or three years, and offenders who are active for more than three years. Unlike figure 3.2, the area under each of the three curves is equal (because I have scaled by level of activity) and we can see the temporal focus of activity is very similar across all three groups.

The high intensity offenders appear to show more activity later on (the curve is higher above age 20). This may be what "life-course" persistence looks like relative to adolescent limited activity. Unfortunately, despite a number of efforts to measure this apparent correlation at no point have I found a significant measure. This may be an area calling for further research.

In figure 3.4 I do the same for λ, grouping individuals by number of offenses.

In figures 3.5 and 3.6I do the same for the Wolfgang arrest data. The pattern is similar, and the similarity in temporal distribution in figure 3.6 is very striking.

In figure 3.7, 3.8, 3.9 and 3.10 I do the same for the NLSY data. The pattern is the same as in the earlier graphs.

Scaling in this way allows us to see how the relative probability of acting changes over the life-course. As can be seen from the figures, when we look at relative probability, the spiky offenders who are active for fewer than three years are very similar to the chronic offenders who are active for more than three years.

As in the earlier figures, there is an apparent difference, as predicted by Moffitt (1993), chronic offenders are more likely to persist in their activity over time (even when controlling for overall activity). However, this difference seems to be more continuous than discrete. The general pattern of correlation between intensity and greater persistence in offending does not show as statistically significant in various techniques of measurement I have tried; continuing efforts to define the pattern in a way that shows statistical significance may be possible with further work.

It is worth noting that when we look at number of acts (figures 3.4, 3.6, 3.8, and 3.10) the similarity in temporal distribution is strikingly similar across intensity levels.

It thus appears that the relative probability of committing a crime is stable across ages 10 to 30; that is to say, if we have two individuals, one with an intensity level of x and the second with an intensity level of $2x$, the odds of the first offending in any period t will tend to be roughly 0.5 of the odds of the second individual offending. In aggregate, there does not seem to be significant difference in patterns.

3.6.2 How Timing Affects GBT Results

In the group-based trajectory method, individuals are grouped based on both intensity and the timing of their activity. Thus, individuals who would be categorized as having similar values of $p(\theta_i, X_i)$ in the model of this paper would be placed in different groups if they offend in different periods.

To understand how individuals who commit a single crime (1/3rd of the Farrington and West offenders) were grouped, I created a small simulated dataset of 11 individuals who were identical except for the period in which they committed a single crime - the first individual committed it in period 1 (age 10-11), the second in period 2 (age 12-13), etc. I used the parameters from Table 6 (Nagin and Land, 1993, p. 349) to estimate the likelihood of their being in each of the three offender groups (High Rate Chronic, Low-Rate Chronic, Adolescent-Limited)[7]. Not surprisingly, none of them were categorized as High Rate Chronics (group A). However, while all the individuals who committed their crime before age 20 were grouped as Adolescent-Limited (group B), all the individuals who committed their crime at age 20 or after were grouped as Low Rate Chronics (group C).

While this is plausible, it is important to note that for roughly half the individuals in these two groups we have only one crime observed. One-third of the offenders in the Farrington and West data committed only one crime, and each of the three offender groups had roughly 1/3rd of the offenders[8] This means that nearly one-sixth of the individuals are grouped as "chronics" based on a single criminal offense.

3.6.3 Comparing Populations

I now present evidence that low-intensity individuals are more like each other, regardless of the timing of their acts, than they are similar to higher intensity individuals who commit crimes at similar ages.

---

[7]Technically, I used slightly different coefficients for the Age and Age^2 variables, dividing the first coefficient by 10 and the second by 100, as this was the only approach that matched the data, the coefficients and their recorded values in tables 3, 4, 5 and 7. I assume that in the original optimization program they divided the age values by 10 so that they were closer in magnitude to the two other variables, TOT and Y_{t-1}. Such rescaling is standard practice in optimization.

[8]The specific breakdown is, in order, 54 offenders, 51 offenders and 40 offenders, see Table 8 Nagin and Land (1993, p. 351).

The first piece of evidence in this regard is presented in table 3.12. In this table, I use table 9 of Nagin and Land (1993, p. 352). This table looks at offenders from the Farrington and West data, divided into three groups by the GBT method: Group A is High-Rate Chronics, Group B is Adolescent-Limited Offenders, and Group C is Low-Rate Chronics. In the model proposed by this paper, Group A would be high intensity, while both Groups B and C would low-intensity. Thus, one way to test the relative strengths of this model vs the GBT method is to compare the population attributes of the three groups - if all three differ from each other, that would support GBT; if B and C are similar but both differ from A, that would support the two-dimensional model.

Table 9 presents data on the percentage of individuals in each of group who possesses a particular attribute (low IQ, say). Using the standard formula for difference-in-means in Bernoulli variables

$$\text{t-statistic} = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

I compute t-statistics for the differences between groups A, B and C. Only one of twenty-four differences between groups B (Adolescent-Limited) and C (Low-Rate Chronics) is significant, and the average of the t-statistic values is close to zero. There do not appear to be significant differences between them. However, one-quarter of the differences between group A and the other two are significant, and the average value of t-statistics is over 1. This suggests that sorting by intensity (high-rate vs low-rate) gets to real differences between individuals, while differences in timing (adolescence vs early adulthood) are at best, much less important.

Additional evidence in this regard comes in table 3.13. In this I compare populations of offenders in the Wolfgang arrest data, sorted by number of offenses (I look at offenders with one or two crimes) and by timing of offenses. The Wolfgang dataset has only a few variables for each offender, so I focus on the measure of socio-economic status as that has the greatest variation. As can be seen, offenders with a single arrest before age 18 are virtually identical in SES to offenders with a single arrest at age 18 or after. However, offenders with two arrests are significantly different. When we focus on individuals with an arrest at a single year (age 15 in this case), the statistical difference between one-arrest and two-arrest offenders disappears, but the difference-in-means between two-arrest offenders and one-arrest offenders is five times greater than the difference between one-arrest offenders with arrests at 15 and 21.

3.6.4 Testing for independence between $\varepsilon(Y_i(t), \omega_i(t)))$ across $[0, T]$ for $\lambda_i = 2$

The graphs provide some evidence that there may be a positive covariance between $p(\theta_i, X_i)$ and either $g(a_i(t))$ or $\varepsilon(Y_i(t), \omega_i(t)))$. In other words, it appears that as intensity $p(\theta_i, X_i)$ increases, activity is slightly more spread out along the life course. However, I am unable to find a statistically significant measure that supports that intuition.

Another question of interest is if there is complete independence between $\varepsilon(Y_i(t), \omega_i(t)))$ across values of $t \in [0, T]$. That is, is there evidence of autocorrelation?

While it is difficult to test this across all data points in all data sets, we can use the χ^2 test to look for it in the Wolfgang data set for individuals with two arrests. For such individuals, we have good reason to assume that all have the same level of intensity $p(\theta_i, X_i)$ (on average) and the same age-crime curve effect $g(a_i(t))$. If the environmental influences are independent, then that implies that we can predict the number of individuals with specific age combinations from the aggregate data. If there is a 15% probability of a $\lambda_i = 2$ individual committing a crime after age 20, and there is independence across periods, then the probability of an individual with $\lambda_i = 2$ committing both crimes after age 20 is $0.15 \times 0.15 = 0.0225$.

We can thus create tables of expected vs. actual crimes for individuals with λ_i and test for significance. In table 3.14 I generate a simple χ^2 table of actual and expected patterns of arrest for the 944 individuals with two arrests in the Wolfgang data set. Using the average probability of any arrest, we can construct a table of expected arrests under a null hypothesis of independence between periods. As can be seen, arrests tend to be bunched in time far more than would be predicted by the null (the same result holds if we construct a similar table with more periods; I use the two-period table for simplicity). Thus we can see strong evidence of the temporal relationship between acts of crime.

The table also allows us to test the GBT prediction of chronic offenders, who are active across the entire period. There are actually far fewer individuals who are active across the entire dataset than would be predicted by chance (the actual number of individuals committing a crime before and after age 20 is 230, much fewer than the 404 predicted under independence). Since this analysis is limited to individuals with two crimes we cannot extrapolate too far, but the lack of "chronics" in this subset would appear to contradict the results of some GBT work, for example D'Unger et al. (1998).

3.6.5 Intensity and Contingent Circumstances

To summarize the analysis of the last few sections, I propose that in analyzing criminal careers we can differentiate individuals by a single stable continuous measure, which I call intensity. This measure is essentially the total number of offenses over an extended period of time (at least five years, ideally a decade or even several decades). Based on previous work it seems likely that this measure can be predicted (albeit imperfectly) on the basis of a range of measures of impatience, impulsiveness, conscientiousness, etc.. This measure would be the best way to summarize an individual criminal career.

There are three other aspects of a career however, that would be necessary to explain the specific pattern of crime of an individual.

The first is simply the well-established age-curve of crime, which in this model can be interpreted as a universal increase in the probability of committing a crime to age 17, followed by the decrease in the probability of committing a crime after that age. Thus the naive probability of an individual committing at least one crime in a particular period (one year, for example) would be determined by the age-curve and scaled by the individual's intensity level.

The second is the temporal relationship between acts of crime for individuals who commit more than one act. The evidence in 3.14 strongly establishes that individuals committing two crimes are significantly more likely to commit them in the same period. This relationship could be modeled in a number of ways, as autocorrelation, or as periods of activity and inactivity. More research is needed to understand this.

The third area is also one where I believe a great deal more research is needed; the contingent circumstances that trigger an individual's decision to commit crime in *this* period and not two periods later. I believe these contingent circumstances are based on real events in the individual's life, in keeping with a range of work on social control issues (Laub et al. (1998), Sampson and Laub (1990)). However, it seems likely that once we have controlled for an individual's intensity level, and the overall shift in criminal activity that occurs over the life course, these circumstances are stochastically independent from each other and from the intensity level. This means that there will not be a second individual-specific stable measure that we can use to predict criminal activity; intensity is the only measure we can use over an extended period of time.

An immediate result of this is that we will need a great deal more longitudinal research on criminal behavior. The fluctuations in activity and the possibility of autocorrelation suggest that it may be helpful to look at behavior on a shorter interval than a year - monthly data could potentially be extremely useful.

3.7 Implications

One of the implicit promises of criminology is that by understanding crime better, we will be able to identify future criminals and prevent them from committing crime. Considering that in most societies brute force is a favored method of crime prevention, it is perhaps not surprising that criminal careers tend to be stochastic and difficult to predict; the easy-to-predict criminals were presumably put out of business long ago.

The analysis presented here has a number of implications.

First, it appears that level of activity is the single most important measure discriminating between different criminals: when we sort by that measure, we see relatively minor differences between groups in the pattern of career, trajectory, etc. but very significant differences in measurable attributes such as drug use, socio-economic status. Several things follow from this. It provides a straightforward explanation for the well-documented fact that age of first offense is a good predictor for lifetime crime; the correlation between these two measures is simply due to the fact that in any year (not just years of early adolescence) a high intensity offender is more likely to be active. Additionally, it suggests that an individual who commits a crime at age 12, but never again, is likely to be quite similar to an individual who commits no crimes in his teenage years, and then one crime at age 25. It further suggests that both of these offenders are quite different from an offender who commits crimes at ages 12, 13 and 14.

Second, an obvious issue that needs to be addressed is what, if anything, can be said *ex ante* about an individual's intensity based purely on behavior up to a certain age. While there is significant value in being able to categorize individuals at age 30, it would be even more helpful to get some sense of possible future behavior by age 17 or 21.

Third, the spikiness means that longitudinal datasets (*pace* Hirschi and Gottfredson (1983)) are essential to track criminal careers. If one-third or more of the population we are interested in are active for only one year, than we need data on as many years as possible. Such data will allow further investigation of both the fluctuation in environmental influences and the apparent autocorrelation in activity in short periods.

Summary of Data Sources			
Data Set	Farrington & West	Wolfgang	NLSY, 1997 Cohort
Basis of Observations	Convictions,)	Arrests	Self-report
Location	London	Philadelphia	United States
Time period	1960's and 70's	1960s, 70's and 80's	1997-2009
Basis of selection	Judged prone to delinquency before age 10	Born in Philadelphia in 1958	Sample of roughly 9,000 teenagers
Ages of individuals	Ages 10 to 24	Birth to age 26	Ages 12-16 to 24-28
Gender	Male only	Male and Female	Male and Female
Crimes	All	All	Theft of items more than $50, Drug dealing

Table 3.1: Descriptions of the three major data sets used in this paper.

	Individuals			Convictions		
Convictions	Number	Percentage	Cumulative Percentage	Counts	Percentage	Cumulative Percentage
1	50	34.2%	34.2%	51	10.7%	10.7%
2	28	19.2%	53.4%	63	13.3%	24.0%
3	22	15.1%	68.5%	66	13.9%	37.9%
4	16	11.0%	79.5%	86	18.1%	56.0%
5	6	4.1%	83.6%	36	7.6%	63.6%
6	6	4.1%	87.7%	29	6.1%	69.7%
7	5	3.4%	91.1%	48	10.1%	79.8%
8	8	5.5%	96.6%	33	6.9%	86.7%
9	3	2.1%	98.6%	33	6.9%	93.7%
10	1	0.7%	99.3%	14	2.9%	96.6%
11		0.0%	99.3%		0.0%	96.6%
12		0.0%	99.3%		0.0%	96.6%
13	1	0.7%	100.0%	16	3.4%	100.0%
Total	146			475		

(265 individuals with no convictions not shown).

Table 3.2: Farrington & West data broken down by number of years active. Notice that one-third of all offenders have a conviction in only one year, and another one third are active in only 2 or 3 years. These offenders are referred to as "spiky" in this paper; Nagin and Land and Barnett et al referred to them as "intermittent"

Career Length (< Years)	Individuals Number	Individuals Percentage	Individuals Cumulative Percentage	Convictions Count	Convictions Percentage	Convictions Cumulative Percentage
1	50	34.2%	34.2%	51	10.7%	10.7%
2	7	4.8%	39.0%	15	3.2%	13.9%
3	11	7.5%	46.6%	25	5.3%	19.2%
4	12	8.2%	54.8%	42	8.8%	28.0%
5	10	6.8%	61.6%	25	5.3%	33.3%
6	7	4.8%	66.4%	27	5.7%	38.9%
7	6	4.1%	70.5%	19	4.0%	42.9%
8	14	9.6%	80.1%	57	12.0%	54.9%
9	5	3.4%	83.6%	21	4.4%	59.4%
10	3	2.1%	85.6%	18	3.8%	63.2%
11	11	7.5%	93.2%	71	14.9%	78.1%
12	4	2.7%	95.9%	32	6.7%	84.8%
13	1	0.7%	96.6%	13	2.7%	87.6%
14	2	1.4%	97.9%	18	3.8%	91.4%
15	3	2.1%	100.0%	41	8.6%	100.0%
Total	146			475		

(265 individuals with no convictions not shown).

Table 3.3: Farrington & West data broken down by length of career (first conviction to last conviction). Notice that due to the “spiky” nature of activity, career lengths are longer than total years of activity. This is true in the other data sets, but career lengths in the self-report data are much closer on average to the count of number of years active.

Years Active	Individuals Number	Percentage	Cumulative Percentage	Arrests Counts	Percentage Percentage	Cumulative
1	2,505	47.0%	47.0%	2,825	15.2%	15.2%
2	1,051	19.7%	66.8%	2562	13.8%	29.1%
3	644	12.1%	78.9%	2,611	14.1%	43.1%
4	411	7.7%	86.6%	2,520	13.6%	56.7%
5	250	4.7%	91.3%	1,931	10.4%	67.2%
6	190	3.6%	94.9%	1,875	10.1%	77.3%
7	111	2.1%	96.9%	1,370	7.4%	84.7%
8	67	1.3%	98.2%	922	5.0%	89.6%
9	47	0.9%	99.1%	830	4.5%	94.1%
10	23	0.4%	99.5%	462	2.5%	96.6%
11	16	0.3%	99.8%	358	1.9%	98.5%
12	6	0.1%	99.9%	157	0.8%	99.4%
13	3	0.1%	100.0%	86	0.5%	99.9%
14	1	0.0%	100.0%	27	0.1%	100.0%
Total	5,325	100		18,536		

(21,835 individuals with no arrests not shown).

Table 3.4: Wolfgang data broken down by number of years active. Notice that nearly one half of all criminals are active (show arrests) in only one year, and another 30% are active in only 2 or 3 years.

Career Length (< Years)	Individuals Number	Percentage	Cumulative Percentage	Arrests Counts	Percentage	Cumulative Percentage
1	2,304	43.3%	43.3%	2,349	12.7%	12.7%
2	427	8.0%	51.3%	1,063	5.7%	18.4%
3	370	6.9%	58.2%	1,061	5.7%	24.1%
4	315	5.9%	64.2%	1,186	6.4%	30.5%
5	289	5.4%	69.6%	1,263	6.8%	37.3%
6	266	5.0%	74.6%	1,235	6.7%	44.0%
7	241	4.5%	79.1%	1,326	7.2%	51.2%
8	222	4.2%	83.3%	1,434	7.7%	58.9%
9	205	3.8%	87.1%	1,373	7.4%	66.3%
10	186	3.5%	90.6%	1,317	7.1%	73.4%
11	160	3.0%	93.6%	1,373	7.4%	80.8%
12	107	2.0%	95.6%	945	5.1%	85.9%
13	96	1.8%	97.4%	1,074	5.8%	91.7%
14	53	1.0%	98.4%	528	2.8%	94.6%
15	42	0.8%	99.2%	526	2.8%	97.4%
16	23	0.4%	99.6%	286	1.5%	98.9%
17	10	0.2%	99.8%	118	0.6%	99.6%
18	7	0.1%	100.0%	57	0.3%	99.9%
19	2	0.0%	100.0%	22	0.1%	100.0%
Total	5,325	100.0%		18,536		

(21,835 individuals with no arrests not shown).

Table 3.5: Wolfgang data broken down by length of career active.

		Individuals			Self-Reported Offenses	
Years of Activity	Number	Percentage	Cumulative Percentage	Counts	Percentage Percentage	Cumulative
1	897	50.4%	50.4%	20,007	10.5%	10.5%
2	396	22.3%	72.7%	30,445	16.0%	26.5%
3	213	12.0%	84.7%	41,349	21.7%	48.1%
4	118	6.6%	91.3%	23,818	12.5%	60.6%
5	66	3.7%	95.1%	19,922	10.4%	71.1%
6	39	2.2%	97.2%	15,364	8.1%	79.1%
7	20	1.1%	98.4%	11,258	5.9%	85.0%
8	17	1.0%	99.3%	18,427	9.7%	94.7%
9	7	0.4%	99.7%	7,374	3.9%	98.5%
10	3	0.2%	99.9%	1,611	0.8%	99.4%
11	1	0.1%	99.9%	341	0.2%	99.6%
12	1	0.1%	100.0%	815	0.4%	100.0%
Total	1,778	100.00%		190,731		

(7,206 individuals who do not report drug-dealing between 1997-2009 not shown.)

Table 3.6: NLSY 1997 Cohort self-reported drug-dealing behavior and offenses summarized by years of activity (all years of survey where drug-dealing is reported).

		Individuals			Self-Reported Offenses	
Career Length (< Years)	Number	Percentage	Cumulative Percentage	Counts	Percentage Percentage	Cumulative
1	897	50.4%	50.4%	20,007	10.5%	10.5%
2	190	10.7%	61.1%	10,349	5.4%	15.9%
3	154	8.7%	69.8%	16,010	8.4%	24.3%
4	115	6.5%	76.3%	13,584	7.1%	31.4%
5	105	5.9%	82.2%	19,554	10.3%	41.7%
6	84	4.7%	86.9%	17,647	9.3%	50.9%
7	65	3.7%	90.6%	16,910	8.9%	59.8%
8	35	2.0%	92.5%	14,650	7.7%	67.5%
9	23	1.3%	93.8%	9,686	5.1%	72.6%
10	36	2.0%	95.8%	12,896	6.8%	79.3%
11	27	1.5%	97.4%	13,961	7.3%	86.6%
12	29	1.6%	99.0%	15,034	7.9%	94.5%
13	18	1.0%	100.0%	10,443	5.5%	100.0%
Total	1,778	100.0%		190,731		

(7,206 individuals who do not report drug-dealing between 1997-2009 not shown.)

Table 3.7: NLSY 1997 Cohort self-reported drug-dealing behavior and offenses summarized length of career (first year from 1997 with reported activity to last year of reported activity up to 2009). Notice that career lengths, while longer than the count for number of years active, are still quite short. This is true for all self-report data (not shown for theft).

Number of Years Active	Individuals Number	Percentage	Cumulative Percentage	Self-Reported Offenses Counts	Percentage Percentage	Cumulative
1	898	71.5%	71.5%	5,785	39.7%	39.7%
2	232	18.5%	90.0%	3,851	26.4%	66.1%
3	74	5.9%	95.9%	2,122	14.6%	80.7%
4	34	2.7%	98.6%	1,621	11.1%	91.8%
5	15	1.2%	99.8%	798	5.5%	97.3%
6	1	0.1%	99.8%	109	0.7%	98.0%
7	1	0.1%	99.9%	165	1.1%	99.2%
8	1	0.1%	100.0%	122	0.8%	100.0%
Total	1,256			14,573		

(7,728 individuals who do not report stealing items worth more than $50 between 1997-2009 not shown.)

Table 3.8: NLSY 1997 Cohort self-report on stealing broken down by number of years active.

Career Length (< Years)	Individuals Number	Percentage	Cumulative Percentage	Self-Reported Offenses Counts	Percentage	Cumulative Percentage
1	898	71.5%	71.5%	5,785	39.7%	39.7%
2	103	8.2%	79.7%	1,841	12.6%	52.3%
3	67	5.3%	85.0%	1,344	9.2%	61.6%
4	51	4.1%	89.1%	1,482	10.2%	71.7%
5	32	2.5%	91.6%	572	3.9%	75.6%
6	35	2.8%	94.4%	861	5.9%	81.6%
7	25	2.0%	96.4%	945	6.5%	88.0%
8	9	0.7%	97.1%	366	2.5%	90.6%
9	8	0.6%	97.8%	418	2.9%	93.4%
10	10	0.8%	98.6%	387	2.7%	96.1%
11	9	0.7%	99.3%	180	1.2%	97.3%
12	7	0.6%	99.8%	206	1.4%	98.7%
13	2	0.2%	100.0%	186	1.3%	100.0%
	1,256	100.00%		14,573		

(7,728 individuals who do not report stealing items worth more than $50 between 1997-2009 not shown.)

Table 3.9: NLSY 1997 Cohort self-report on stealing broken down by length of career.

Number of Rapes	Individuals Reporting	Rapes	Percentage (Low)	Percentage (High)
1	44	44	12%	5%
2	34	68	19%	8%
3	7	21	6%	3%
4	9	36	10%	4%
5	4	20	6%	2%
6	5	30	8%	4%
7	3	21	6%	3%
8	3	24	7%	3%
9 to 50	11	99 (9×11)	27%	
		550 (50×11)		68%
Total (Low)		363		
Total (High)		814		

Table 3.10: Self-admitted rape offense data from Lisak and Miller (2002). Because the data for individuals reporting committing nine or more rapes are grouped, I use both the bottom of the range (9 rapes) and the top (50 rapes) to look at what percentage of rapes are committed by the chronic offenders. Notice that a substantial number of individuals have only committed a few rapes. There is insufficient data in the survey to develop a longitudinal summary of rape offenses, but it seems likely that at least some rapists have only been active for short periods.

Period of conviction	Age at conviction	Grouping	Log Likelihood if Group A	Group B	Group C	Group D
1	10-11	B	-10.10	-4.24	-4.62	-4.72
2	12-13	B	-9.90	-3.27	-4.34	-4.72
3	14-15	B	-9.76	-2.82	-4.11	-4.72
4	16-17	B	-9.68	-2.89	-3.92	-4.72
5	18-19	B	-9.66	-3.48	-3.78	-4.72
6	20-21	C	-9.71	-4.60	-3.69	-4.72
7	22-23	C	-9.82	-6.23	-3.64	-4.72
8	24-25	C	-10.00	-8.39	-3.64	-4.72
9	26-27	C	-10.24	-11.07	-3.69	-4.72
10	28-29	C	-10.54	-14.27	-3.79	-4.72
11	30	C	-10.90	-17.94	-3.93	-4.72

Table 3.11: As an exercise in understanding the Nagin and Land GBT methodology as implemented on the Farrington and West data, I constructed a pseudo-dataset of 11 individuals, each with a TOT value of 0.25 (roughly the sample mean), each with a single offense in one of the 11 periods. I then used the parameters of Nagin and Land's second stage model to generate a log likelihood of observing that offense pattern under the four parameterizations, and then assign each of these pseudo individuals to a different group. Those whose offense occurs before age 20 are placed in group B, Adolescent-Limited, while those whose offense occurs from age 20 onwards are place in group C, Low-Rate Chronics. This seems arbitrary, and the lack of significant difference in population means (see table 3.12) supports this. Moreover, one-third of all offenders in the Farrington and West data have committed only one offense. This is roughly 1/2 of all offenders who are placed in either group B or group C, meaning that the group division is in very substantial part being constructed on these "singleton" offenders.

	Group A to C	Group A to B	Group B to C
Low IQ	-1.44	-0.13	-1.31
Poor Child Rearing	-0.23	-0.40	0.14
Daring	-0.42	-0.86	0.38
Criminal Parents	-0.06	1.21	-1.17
Troublesome	1.73	1.30	0.48
Low Parent Interest Education	0.89	1.26	-0.29
Poor Supervision	-1.11	-0.63	-0.52
Delinquent Siblings	3.17**	2.54*	0.70
Lacks Concentration	3.02**	2.75**	0.40
Conduct Disorder	0.96	1.05	-0.03
Physical Neglect	-0.24	-0.14	-0.11
Poor Housing	-0.59	-0.16	-0.43
Low peer popularity	-0.03	1.52	-1.39
Poor Junior School Performance	0.75	2.37*	-1.41
Frequently Truant	1.14	1.14	0.05
Disobedient	1.74	1.42	0.38
Lies	2.52*	2.22*	0.38
Hostile to Police (12 to 15)	1.93	0.93	1.02
Hostile to Police (16 to 19)	0.50	0.99	-0.42
Low Job Stability	2.98**	1.91	1.10
Heavy Beer Drinking	2.01*	1.97*	0.15
Marijuana	3.85***	2.13*	1.66
Frequent Gambling	1.18	0.82	0.40
High Pro-Aggressive Attitude	1.64	1.22	0.47
High self-reported violence	3.58***	1.12	2.38*
Average of t-statistic values	1.18	1.10	0.12
Number that are significant	7	6	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.12: Using the standard formula for difference in means in Bernoulli variables

$$\text{t-statistic} = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

I compute t-statistics for the differences between groups A, B and C in the attribute measures provided in table 9 of Nagin and Land (1993, p. 352). Notice that only one of twenty-four differences between groups B (Adolescent-Limited) and C (Low-Rate Chronics) is significant, and the average of the t-statistic values is close zero, while one-quarter of the differences between group A and the other two are significant, and the average value of t-statistics is over 1. This suggests that sorting by intensity (high-rate vs low-rate) gets to real differences between individuals, while differences in timing (adolescence vs early adulthood) are not important.

	Single Arrest Before 18	Single Arrest 18 and After	Single Arrest	Two arrests
N	1,433	835	2,268	944
Socio-economic Status	-0.253	-0.256	-0.254	-0.406
(Std. Dev)	0.954	0.943	0.950	0.929
Difference	0.003		0.152	
SE	0.041		0.036	
t-statistic	0.064		4.195***	

	Single Arrest Age 15	Two arrests One at Age 15	Single Arrest Age 15	Single Arrest Age 21
N	291	186	291	88
Socio-economic Status	-0.207	-0.299	-0.207	-0.187
(Std. Dev)	0.936	0.984	0.936	0.891
Difference	0.092		-0.020	
SE	0.091		0.110	
t-statistic	1.012		-0.179	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.13: Working with the Wolfgang data set, I compare the mean value of the socio-economic status measure for populations of criminals categorized by (a) number of crimes (1 or 2) and (b) timing of offenses. As can be seen, offenders who commit a single crime before age 18 are virtually identical in SES measure to offenders who commit a single crime at 18 or after, while offenders who commit two crimes show a strong statistically significant difference. Focusing on offenders who commit a crime at age 15 there is not a statistically significant difference between offenders who commit one crime or two, but the difference in SES is much less between single crime offenders than between single and double crime offenders.

χ^2 test of independence across periods, $\lambda_i = 2$

	Both crimes before age 20	One crime before, one crime after	Both crimes afte
Expected Individuals	90.94	404.12	448.94
Actual Individuals	178	230	536
N	944		
χ^2 Value	175.24		
degrees of freedom	3		
p-value	0.00		

Table 3.14: Using the Wolfgang dataset we can test whether or not the probability of criminal acts are independent across periods. Focusing on the 944 individuals with 2 arrests, and using the average probability of any arrest, we can construct a table of expected arrests under a null hypothesis of independence between periods. As can be seen, arrests tend to be bunched in time far more than would be predicted by the null (the same result holds if we construct a similar table with more periods). One results is that there are actually far fewer individuals who are active across the entire dataset than would be predicted by chance (the actual individuals committing a crime before and after age 20 is 230, much fewer than the 404 predicted under independence). Since this analysis is limited to individuals with two crimes we cannot extrapolate too far, but the lack of "chronics" would appear to contradict the results of some GBT work, for example D'Unger et al. (1998).

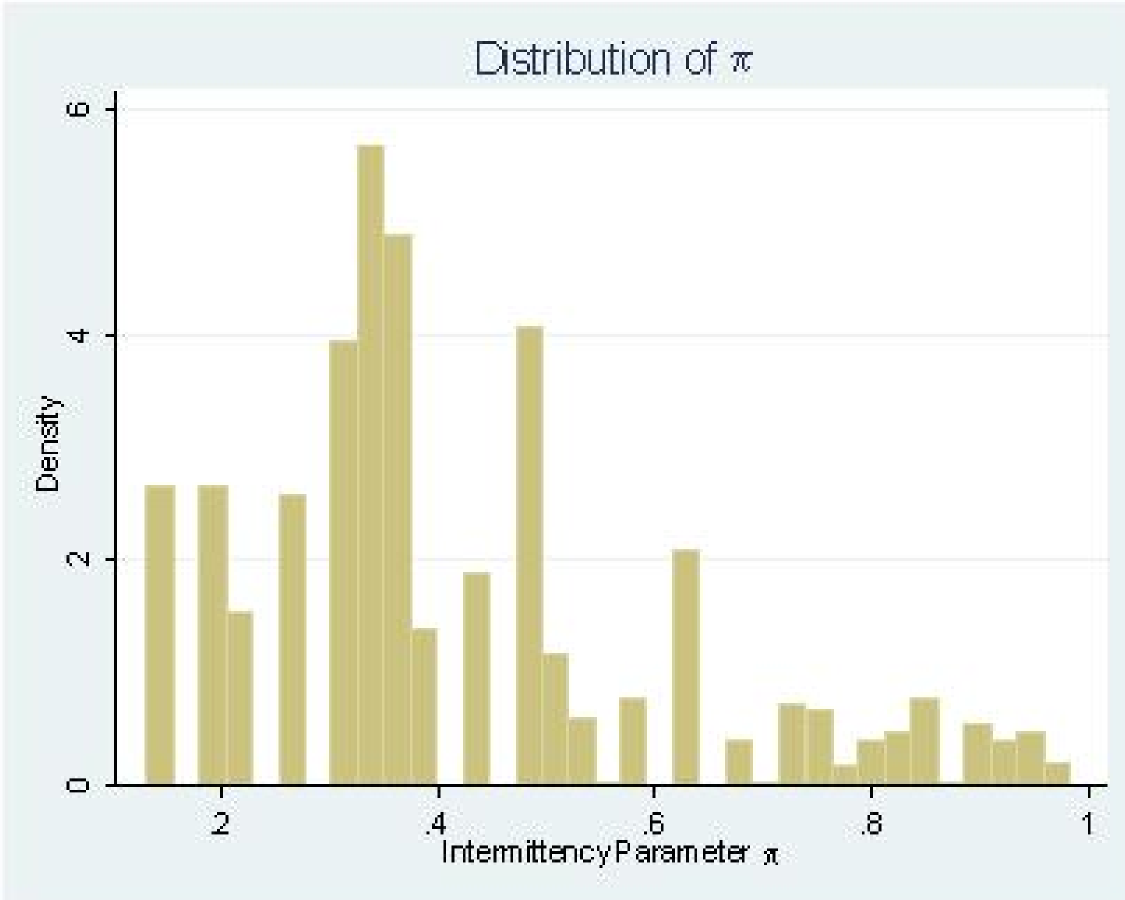

Figure 3.1: A histogram of values for π, the intermittency parameter, using the values from Table 6 (Barnett et al., 1989, p. 349). The distribution seems to be scattered, without a clear central mass, suggesting that it is not estimating an underlying population variable.

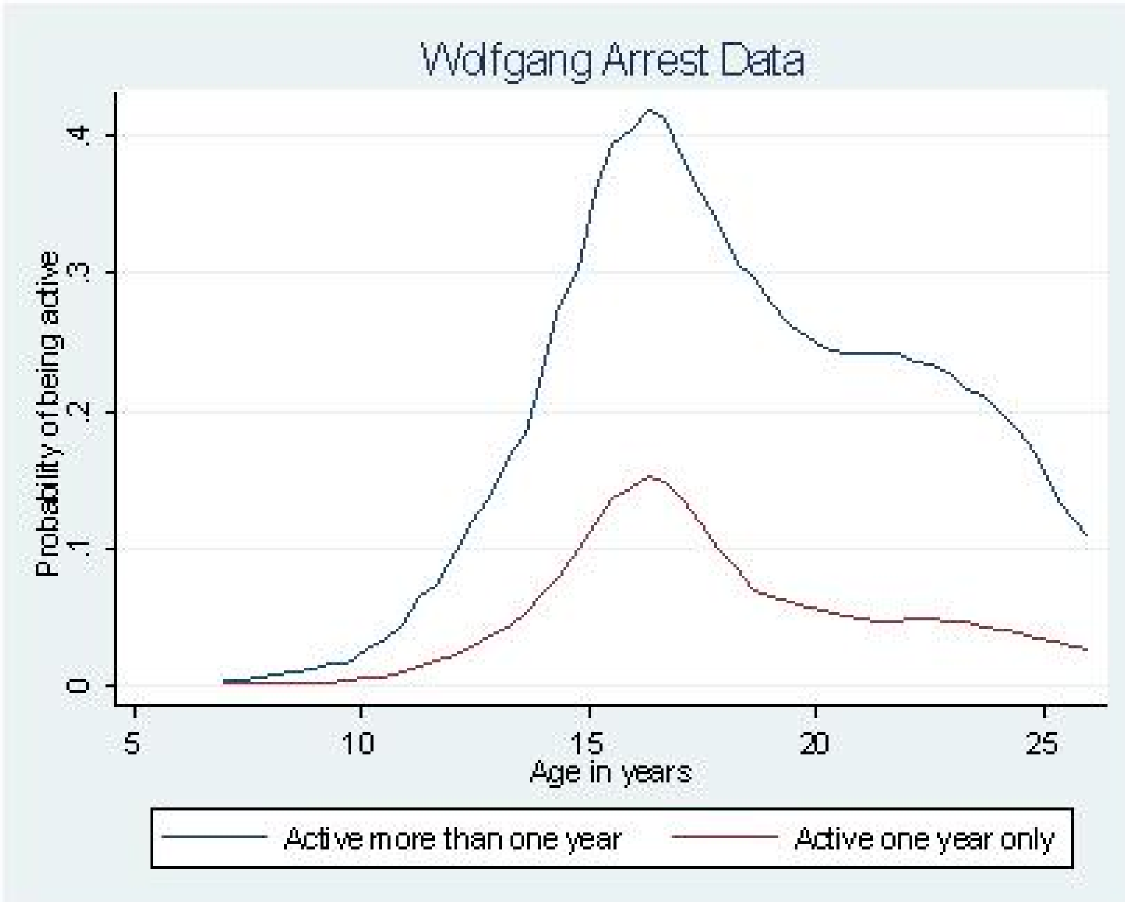

Figure 3.2: A kernel regression of the probability of a male being "active" against age. Notice that while the higher intensity criminals are more likely to be active in any period, the aggregate pattern of both groups changes in the same way over time (increasing at a similar rate to age 17, decreasing in a similar way thereafter).

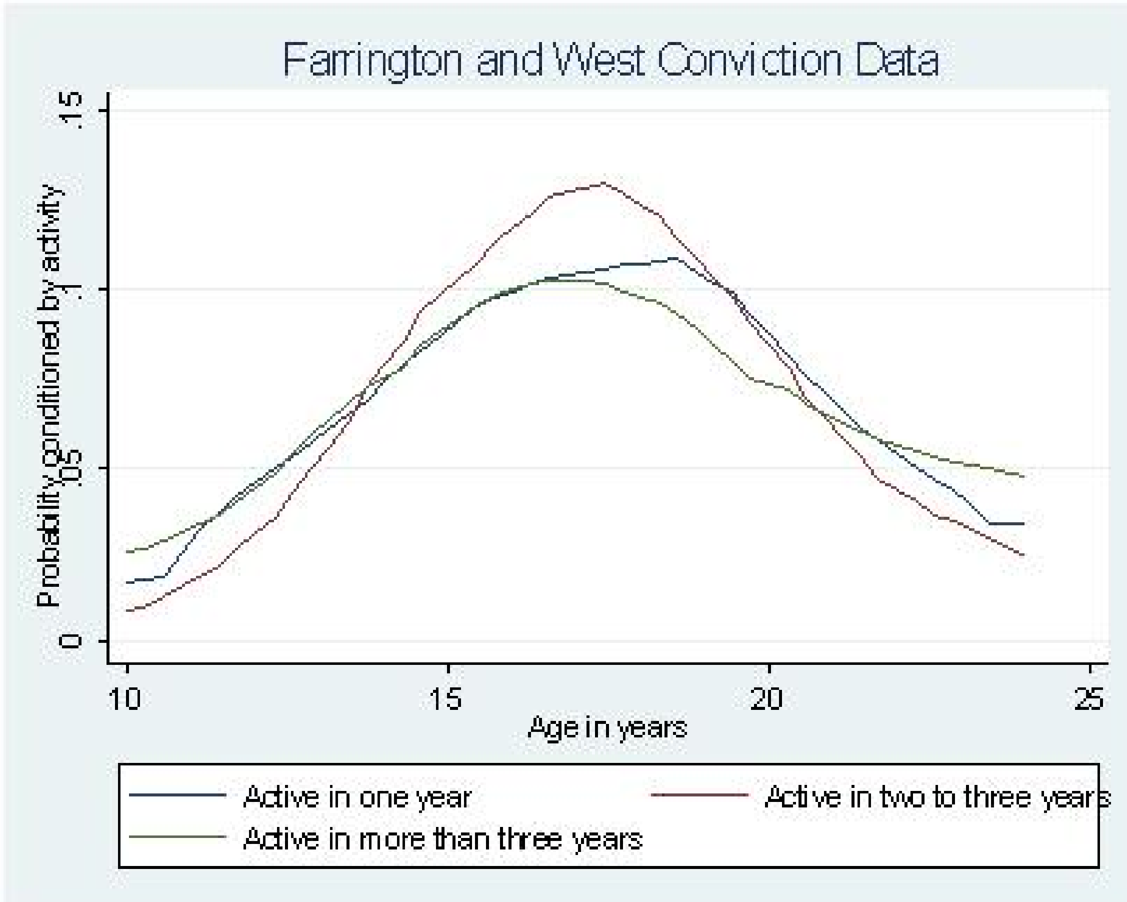

Figure 3.3: A kernel regression of the scaled probability of an individual being "active" against age. An individual is defined as "active" if he receives one or more convictions during a year. This is then scaled by total number of years active (i.e., an individual active in one year would be coded 1.0 in that year, 0 in all others; an individual active in two years would be coded 0.5 in both years, 0 in all other years, etc.). Data from Farrington and West. Notice that all three groups show very similar patterns; chronic offenders with 3 or more years with convictions are more likely to be active after age 17-18.

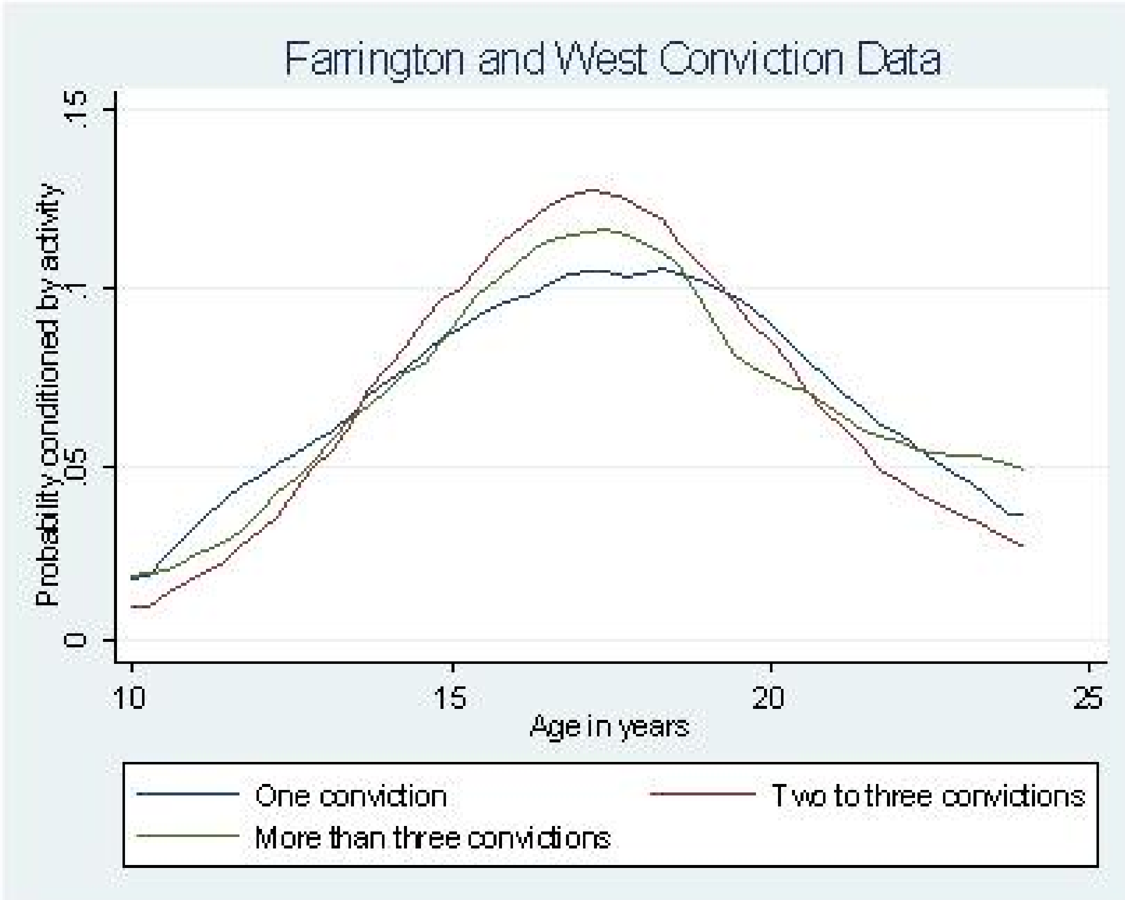

Figure 3.4: A kernel regression of the scaled probability of an individual being "active" against age. While the previous figure 3.3 focused on whether or not an individual was active in a single year, this figure focuses on measures of lambda. (Thus, for an individual who recieves three convictions in a single year, that year would be coded 1.0, the remainder zero, but if two convictions occurred in one year and the third the year after, these would be coded 0.66 and 0.33 respectively). Data from Farrington and West. Notice that all three groups show very similar patterns; chronic offenders with 3 or more convictions are more likely to be active after age 17-18.

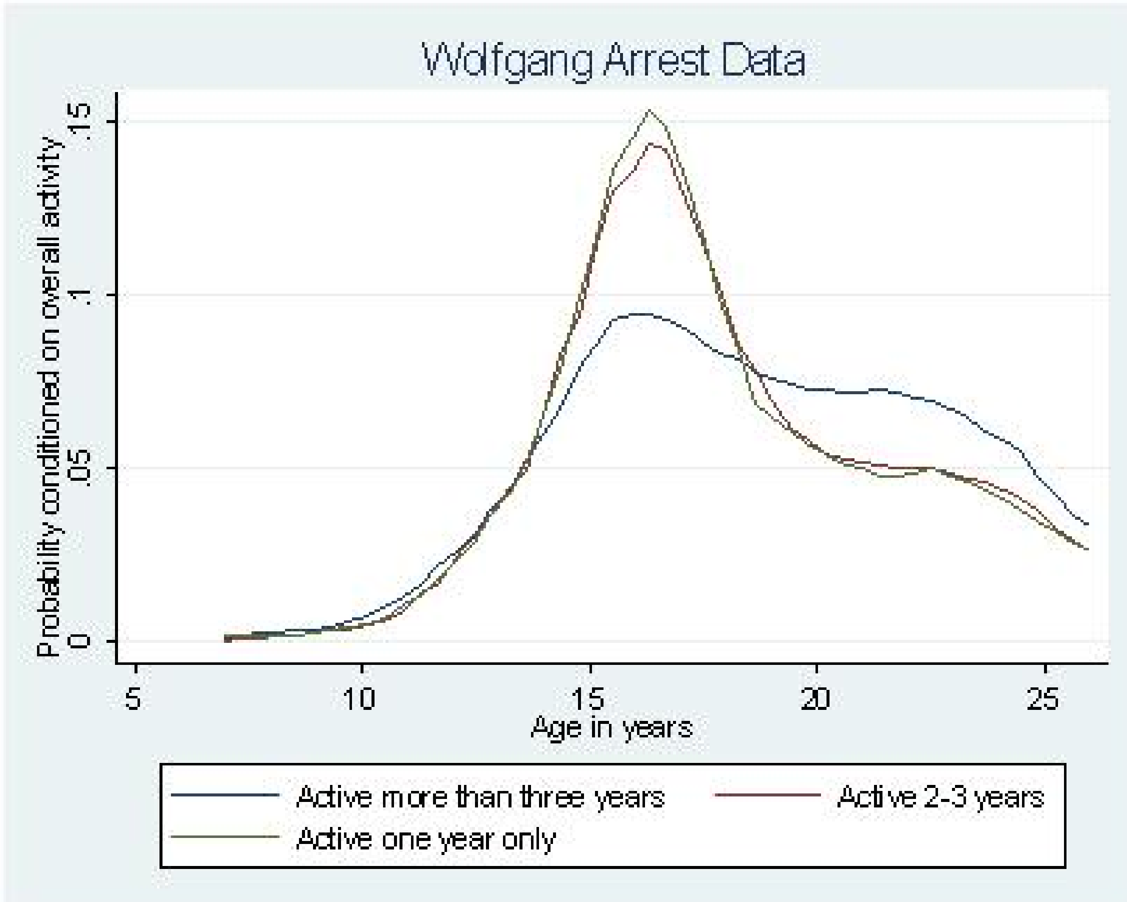

Figure 3.5: A kernel regression of the scaled probability of a male being "active" against age. An individual is defined as "active" if he is arrested one or more times during a year. This is then scaled by total number of years active (i.e., an individual active in one year would be coded 1.0 in that year, 0 in all others; an individual active in two years would be coded 0.5 in both years, 0 in all other years, etc.). Data from Wolfgang 1958 cohort. Notice that all three groups show very similar patterns; chronic offenders with 3 or more years of arrests are more likely to be active after age 17-18.

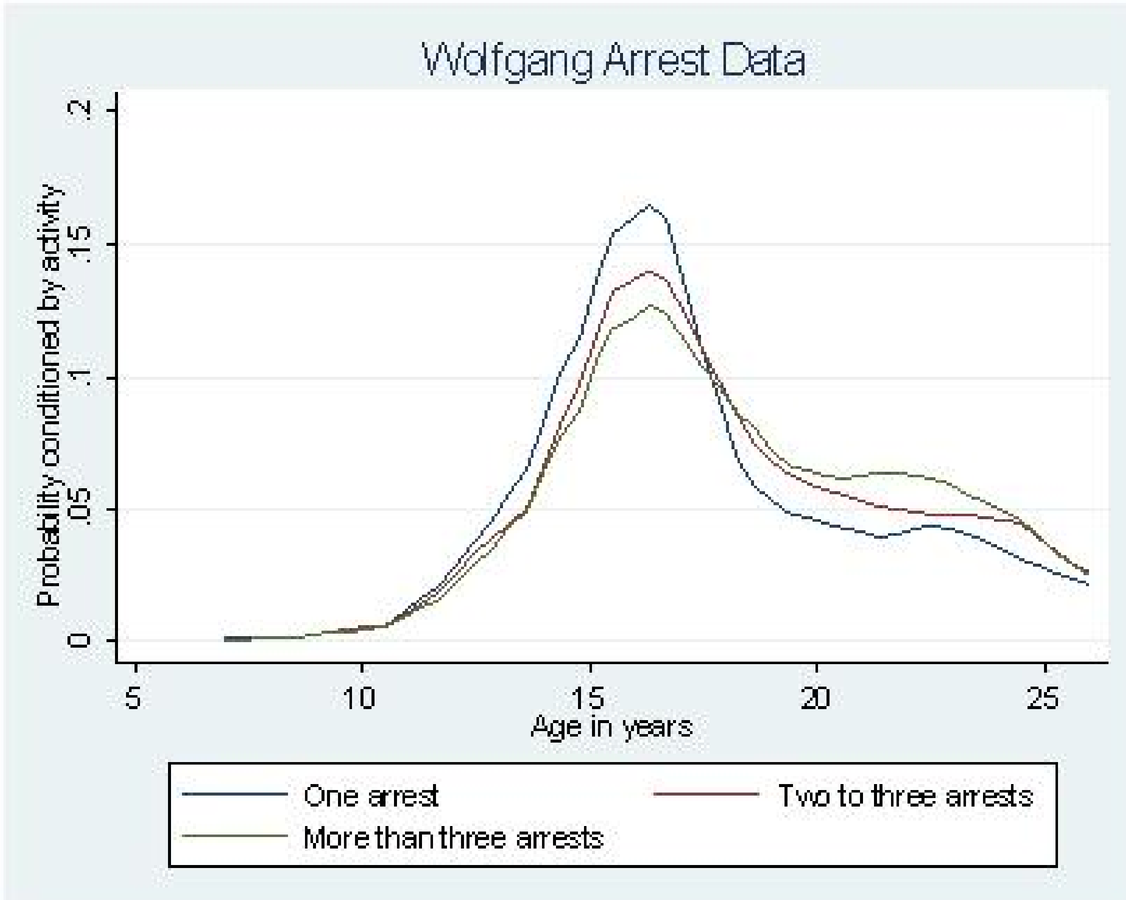

Figure 3.6: A kernel regression of the scaled probability of an individual being "active" against age. While the previous figure 3.5 focused on whether or not an individual was active in a single year, this figure focuses on measures of lambda. (Thus, for an individual who is arrested three times in a single year, that year would be coded 1.0, the remainder zero, but if two arrests occurred in one year and the third the year after, the two years would be coded 0.66 and 0.33 respectively, and all other years coded zero). Data from Wolfgang 1958 cohort. Notice that all three groups show very similar patterns; chronic offenders with 3 or more arrests are more likely to be active after age 17-18.

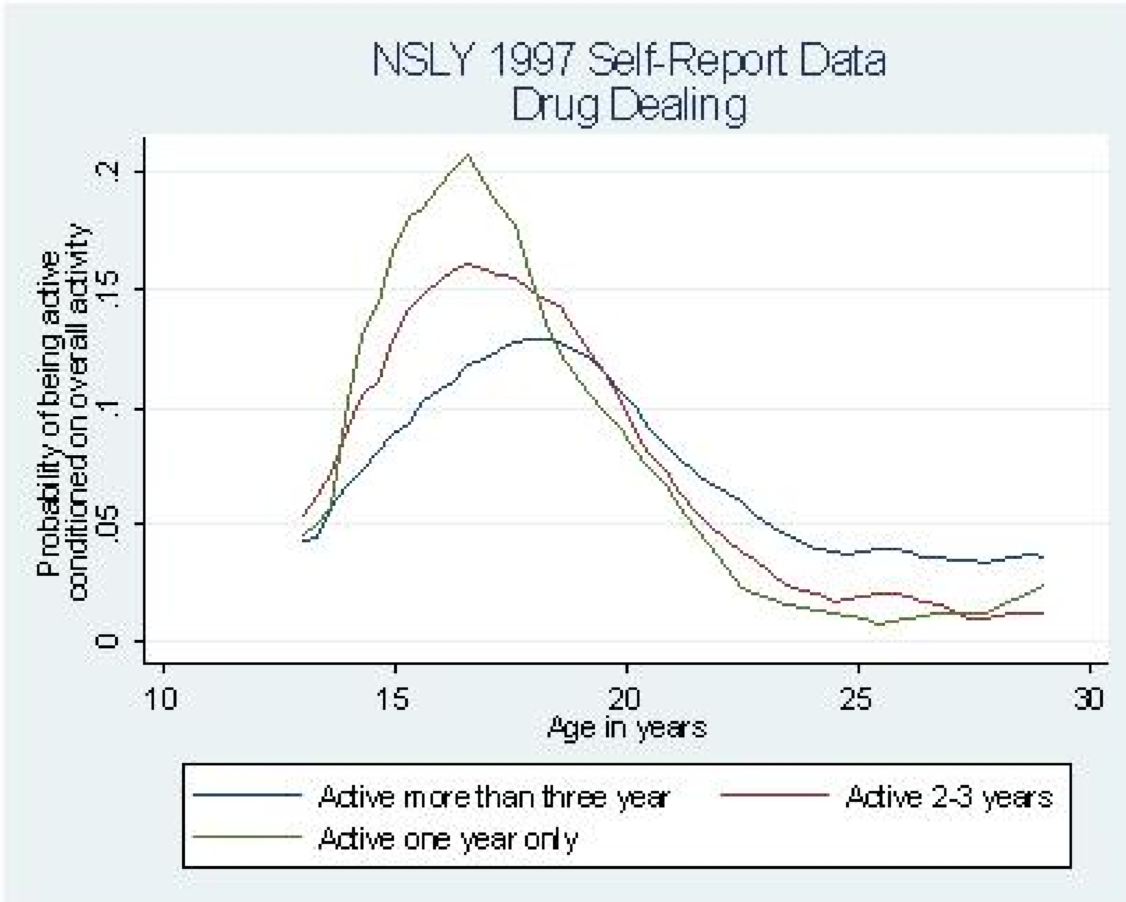

Figure 3.7: A kernel regression of the scaled probability of a self-reported drug dealer being "active" against age. An individual is defined as "active" if he reports one or more drug transactions during a year. This is then scaled by total number of years active (i.e., an individual active in one year would be coded 1.0 in that year, 0 in all others; an individual active in two years would be coded 0.5 in both years, 0 in all other years, etc.). Data from NLSY. Notice that all three groups show very similar patterns; chronic offenders with 3 or more years of activity are more likely to be active after age 17-18.

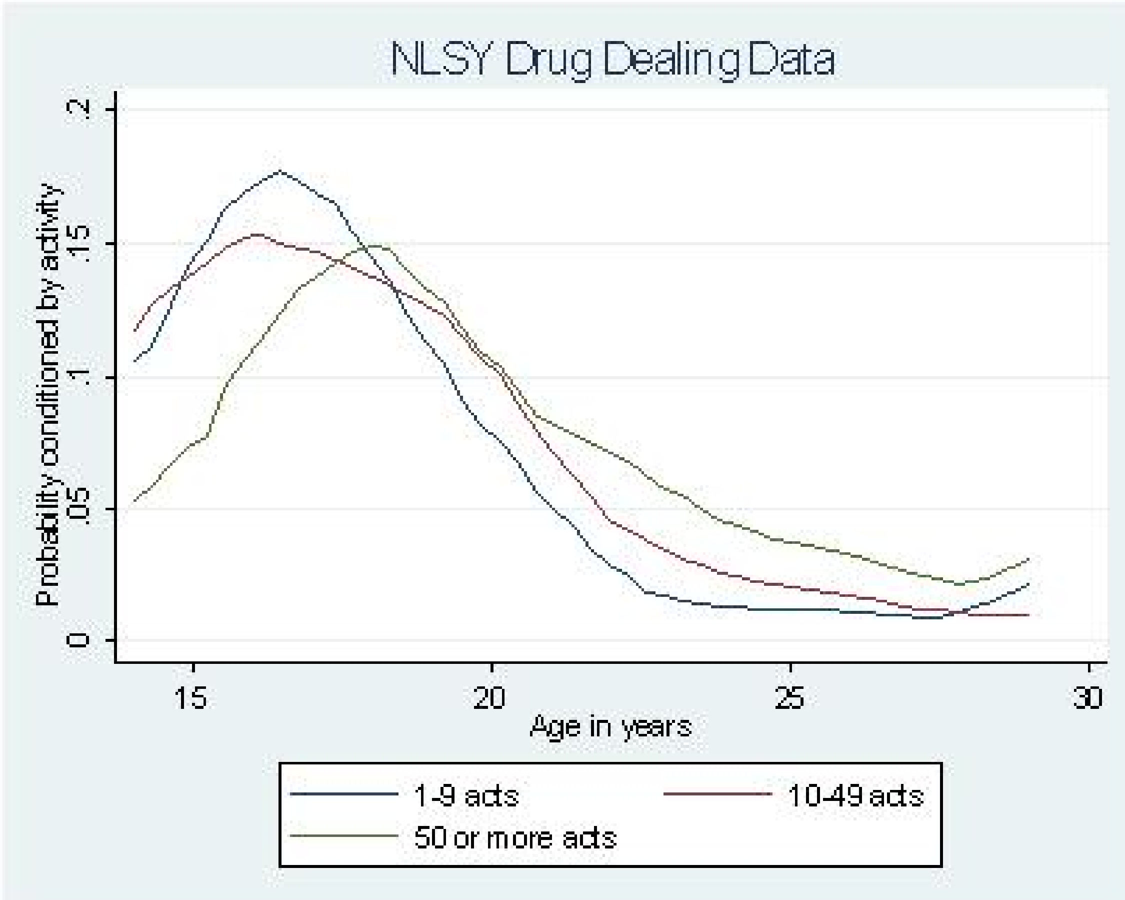

Figure 3.8: A kernel regression of the scaled probability of an individual being "active" against age. While the previous figure 3.7 focused on whether or not an individual was active in a single year, this figure focuses on measures of lambda. (Thus, for an individual who reports three acts in a single year, that year would be coded 1.0, the remainder zero, but if two acts occurred in one year and the third the year after, the two years would be coded 0.66 and 0.33 respectively, and all other years coded zero). Data from NLSY. Notice that all three groups show very similar patterns; chronic offenders with 50 or more reported deals are more likely to be active after age 17-18.

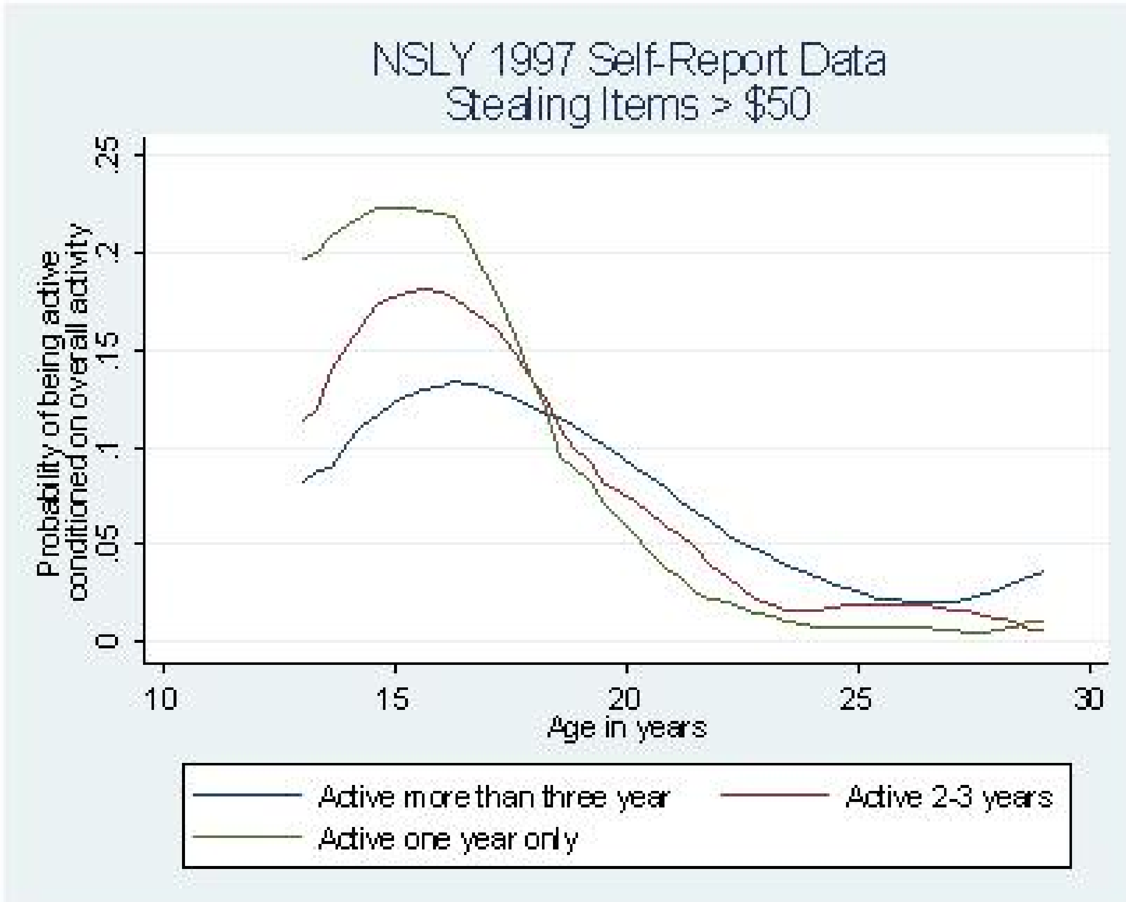

Figure 3.9: A kernel regression of the scaled probability of a self-reported thief being "active" against age. An individual is defined as "active" if he reports stealing one or more items worth more than $50 during a year. This is then scaled by total number of years active (i.e., an individual active in one year would be coded 1.0 in that year, 0 in all others; an individual active in two years would be coded 0.5 in both years, 0 in all other years, etc.). Data from NLSY. Notice that all three groups show very similar patterns; chronic offenders with 3 or more years of activity are more likely to be active after age 17-18.

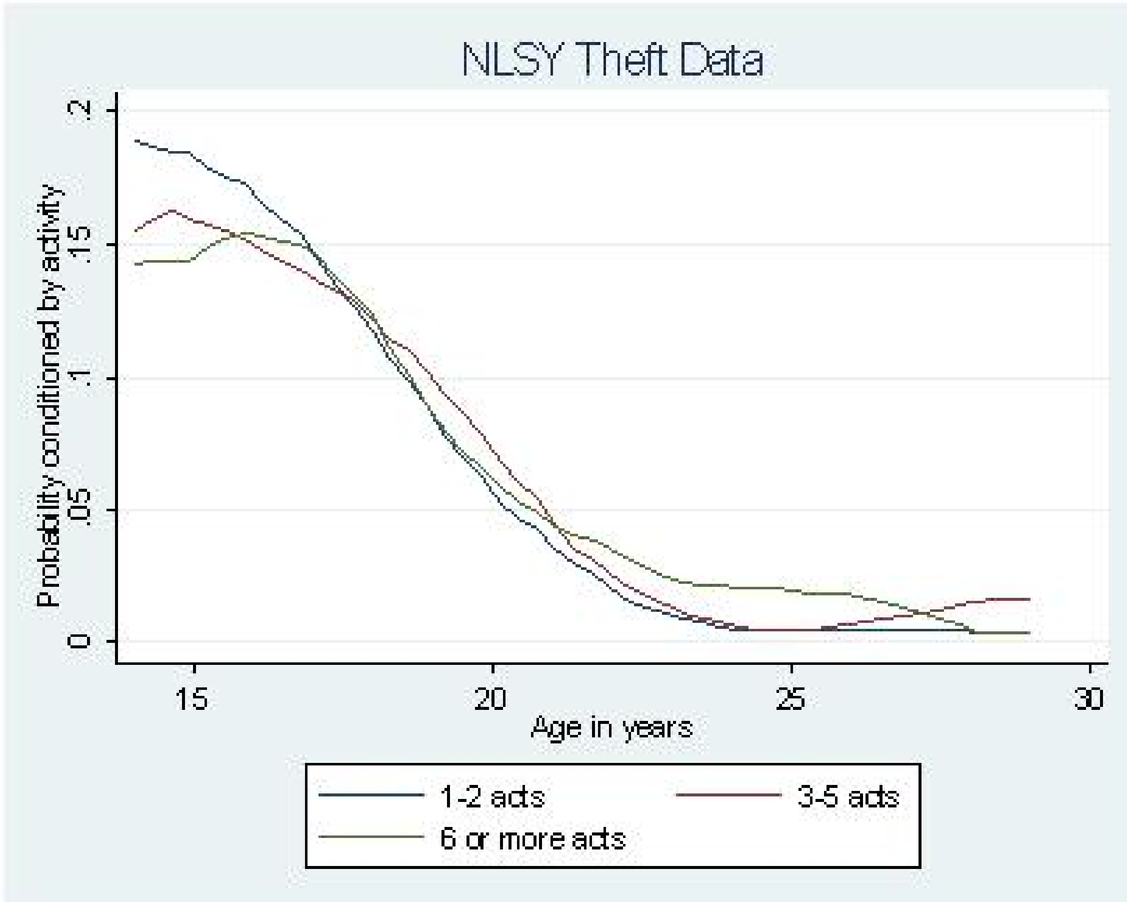

Figure 3.10: A kernel regression of the scaled probability of an individual being "active" against age. While the previous figure 3.9 focused on whether or not an individual was active in a single year, this figure focuses on measures of lambda. (Thus, for an individual who reports three acts in a single year, that year would be coded 1.0, the remainder zero, but if two acts occurred in one year and the third the year after, the two years would be coded 0.66 and 0.33 respectively, and all other years coded zero). Data from NLSY. Notice that all three groups show very similar patterns; chronic offenders with 6 or more acts of theft are slightly more likely to be active after age 17-18.

Chapter 4

The Thief's Wages: Understanding Theft Within A Larger Human Capital Process

4.1 Introduction

Since at least Becker (1968) a standard explanation for criminal behavior is that illegal employment is a substitute for legal employment. In its simplest form, the theory argues that the individual chooses between a legal job and an illegal activity, comparing the relative "wages" and overall costs of the two activities. Thus, according to this theory, when the benefits for legal work fall, illegal work becomes more attractive and we should see more theft, drug-dealing, etc.

While straightforward in theory, the empirical evidence supporting this view is not overwhelming. Wages of criminals are not consistently less than noncriminals, either before and after conviction. Returns from criminal behavior appear to be low on average, but with very high variance across individuals and periods.[1]

An alternative explanation for criminal behavior is that criminals are more present-oriented and less patient than non-criminals. While objective costs and benefits play some role, the major difference is that criminals prioritize the short-term benefits of criminal behavior over the long-term costs and risks.

These two broad schools explaining criminal behavior - labor substitution vs impatience - are not completely exclusive and in fact both presumably play some role. It is almost

[1] I review the literature below.

certain that a decrease in wages or employment opportunities will have some influence on the margin on the decision to steal. Economic theory and common experience would similarly predict that individuals vary in their time preferences and this would make some more likely to steal than others, even if the labor market opportunities available to both groups were identical.

However, when we look at which of the two issues plays a *dominant* role in causing theft, this leads to two very different predictions in human capital development. If labor substitution is the key causal factor, this would predict that on average criminals are significantly lower in human capital than non-criminals, especially during the late teens and early twenties where criminal activity is at its highest. Specifically, we would predict that if we can see differences in wages and other measures of labor market success between different groups in the data, we would see differences between thieves and non-thieves. If impatience is the key causal factor, this would predict that criminals, whatever their relative human capital levels, would invest significantly less in human capital than non-criminals. Specifically, they would be less likely to be enrolled in school, when enrolled in school they would be less successful, and their wages would slip relative to the wages of non-thieves.

In this paper I develop a model that integrates human capital development and criminal behavior choices (specifically theft) and compare the actual patterns of human capital development in the NLSY 1997 with those predicted by the model. The development of the wages and weeks worked of thieves strongly suggests they have roughly equivalent human capital but are much less patient than non-thieves. The wages and hours worked of thieves are close to the wages and hours worked of non-thieves, and some researchers find slightly higher levels in early life (Nagin and Waldfogel, 1995). The lack of a consistent difference in wages over the teens and twenties is in strong contrast with social background or race, where groups that might be predicted to have lower human capital show significantly lower wages and lower weeks worked.

Additionally, thieves show less investment in human capital by a number of measures.

They show less stability over jobs (on average having more employers) and are more likely to repeat grades, even when controlling for a range of measures of socioeconomic status and ability.

I thus conclude that the difference in impatience is far more important than this-period human capital levels.

The paper proceeds as follows. In section 2 I review the literature on earnings from crime and the honest earnings of criminals, as well as the literature on human capital. In section 3 I develop and solve a model that links the decision to steal with the decision to invest in human capital. I look at what the model predicts with range of parameterizations. In section 4 I summarize some of the broad patterns in the NLSY 1997 data and then look at the development of wages and weeks worked among thieves and non-thieves. I regress a number of other measures of human capital development on covariates, and find that theft behavior is generally closely linked to underinvestment in human capital, even controlling for basic measures of human capital. In section 5 I summarize the results.

4.2 Literature

4.2.1 Crime as Substitute for Legal Work

While Becker laid the groundwork, Richard Freeman is perhaps the researcher of the past few decades who has done the most to develop the theory of crime as a substitute for legal work, and what that means in our understanding of both crime and work (Freeman (1999), Freeman (1991)). Recent papers to show a link using individual-level data on actual wages are Grogger (1998) and Lochner (2004). A number of papers find interesting patterns at an aggregate level, including Gould et al. (2002).

Criminologists have been more skeptical of the idea, and a number of researchers in that field find little or no evidence that crime is a rational choice. As Wilson and Abrahamse

(1992) argue, while a single crime can be a rational choice, given a low probability of capture ("The wonder is that more people don't steal.") a criminal career seems not to pay off at all. One of the dominant theories of crime, outlined in Gottfredson and Hirschi (1990) is in fact based on the conjecture that criminals simply have less self-control. Even a number of economists have found evidence that disagrees. Levitt and Venkatesh (2000) examine the revenues and wages provided by drug-dealing in a gang in Chicago and conclude "it is difficult (but not impossible) to reconcile the behavior of the gang members with an optimizing economic model without assuming nonstandard preferences or bringing in social/nonpecuniary benefits of gang participation." Lee and McCrary (2005) look at the deterrent power of prison sentences and conclude "criminal behavior – at least for the kinds of crimes that we focus on – could be thought of as the consequence of a self-control problem and a taste for immediate gratification."

Some researchers have found that criminals actually have slightly *higher* wages than non-criminals. Specifically, Nagin and Waldfogel (1995) find that conviction leads to greater instability but also higher pay. They quote West and Farrington (1977) who noted a "tendency for delinquents and especially recidivists to take up laboring or unskilled occupations offering relatively high rates of pay for beginners, but with relatively few prospects of long-term advancement. Non-delinquents were more likely to defer immediate material rewards for the sake of obtaining apprenticeships or training for skilled work."

There is no consensus on whether or not there is damage to employment wages from a criminal record, and what that damage might be, if it existed. Freeman (1991) finds that incarceration reduces weeks worked between 20 and 8 weeks per year, and reduces the probability of work by 15 to 30 percent. Grogger (1995) and Nagin and Waldfogel (1995) (cited above) are much more cautious; both find limited effects on wages and employment (in some cases positive).

4.2.2 Human Capital and Wage Development

While Friedman and Kuznets (1954) may be the original development of the logic of human capital, Becker (1962) seems to have been the critical article to put all the pieces in place. Additional work was done by Mincer (1974) and Ben-Porath (1967).

James Heckman and his students have done enormous work in this area in the past few decades, including (to name a few contributions) Cunha and Heckman (2007), Cunha et al. (2006) and Cunha et al. (2010).

The paper that most resembles this one is Lochner (2004), in that both use a human capital model based on the Ben-Porath model and then test this model against NLSY data. The major difference in approach is that the model in this paper is much more stripped down. The simplification of the model is a trade-off; as a negative, it ignores a number of real-world issues, but as a plus, allow for a more comprehensive analysis of optimal behavior. Specifically, my model satisfies both Kuhn-Tucker necessity and Kuhn-Tucker sufficiency allowing me to identify a single optimum. I believe this is a justifiable move towards the lamppost and away from the keys. I use the NLSY 1997 which provides more detailed data as opposed to the NLSY 1979 that Lochner uses. The data on returns from theft lead to a different approach to modeling crime - I model a theft as a “negative lottery” with a strong stochastic element. Finally, I come to different conclusions, arguing that the discount rate is crucial to understanding both theft and human capital development, while Lochner emphasizes differences in initial human capital levels as crucial.

4.3 Model

The model focuses on understanding the decision to steal in the context of overall career choices and lifetime income. Thus, one component of the model is a human capital investment problem. The other component of the model is a fixed sequence of decisions to steal or not to steal. The decision to steal is modeled as a “negative lottery,” where theft is largely about the trade-off between immediate returns and long-term risks.

Partly because it fits the data well, and partly for reasons of mathematical simplicity, the human capital investment decision and the decision to steal are not incorporated in a single equation, but instead are kept separate and affect each other indirectly. Each agent lives for T periods. At the beginning of each period agents set their human capital investment levels, based on their parameters (human capital level, discount rate, human capital production parameters, etc). In one period, the first period, they also have the opportunity to steal after setting their human capital investment level; they are exposed to a sequence of criminal opportunities that they can make the binary choice to take or pass on. If they choose to steal, there is a risk of punishment which in addition to direct disutility, inflicts a negative shock to human capital levels and production choices. If their human capital is "tarnished", then in the next period (period 2) they will need to reoptimize their human capital investment levels accordingly.[2]

4.3.1 Human Capital Component

There is a population of N agents, indexed $i \in \{1, \ldots, N\}$ who are maximizing utility over T periods. In each period t each agent chooses what percentage of their human capital to hold back from the market as "investment", which I denote with the variable $x_{it} \in [0, 1]$. The value of x_{it} ranges from 1 for a full time student with no job to 0 for a full-time worker at a job receiving no training, and would be somewhere in the range $(0, 1)$ if i is receiving some wage compensation but is also receiving some training, formal or informal. The problem in period 1 is thus formally:

$$\max_{\{x_{it}\}_1^T} \sum_{t=1}^{T} E(\beta_i^{(t-1)}(1 - x_{it})K_{it})$$

$$\text{s.t. } K_{it+1} = d_{0i}(x_{it}K_{it})^{d_{1i}} + K_{it}$$

$$K_{i0} = \hat{K}_i$$

[2]The decision to allow theft in only a single period both fits the data (the vast majority of thieves are active for a short period in their late teens or early twenties) and also simplifies the mathematical analysis. A more comprehensive model would allow the opportunity to steal in each period, as in the first chapter.

That is, the agent i needs to find an optimal sequence of investment decisions $\{x_{it}\}_1^T$ given the parameter values and his initial level of human capital $\hat{K}_i$.

There are several key points to be made about this problem.

First, this is a simplified variant of the Ben-Porath human capital model, which is the standard dynamic optimization capital accumulation model with the wrinkle that the capital is not consumed.

Second, the model does not include a depreciation parameter; this decision is made for simplicity, and because for the population we will be focusing on, youths age 12 to 24, all evidence suggests that human capital growth swamps human capital decay and it is difficult to separate any decline in human capital from a general failure to invest in human capital.

At the beginning of every period $\hat{t}$, agent i can reassess his level of human capital $K_{i\hat{t}}$ and reset his planned sequence of investment $\{x_{it}\}_{\hat{t}}^T$.

4.3.2 Criminal Opportunities Component

The second component of the model involves criminal activity. There is a single place when individuals have the opportunity to steal; in between the investment decisions of period 1 and period 2 the agent is sequentially offered S opportunities to steal. For the remainder of the agents life, period 2 through T, there will be no further theft opportunities.

Each opportunity can be thought of as a “negative” lottery: a standard lottery asks you to pay out a sum of money in exchange for the possibility of winning a positive sum in the future. In contrast, these theft opportunities offer the agent a positive sum of money in exchange for the possibility of a future loss. The cash value of the opportunity is denoted r_{is} and has a random distribution $f(r_{is};\theta)$ where $f()$ has support $(0,\infty)$ and is decreasing in r_{is}. To simplify matters, I assume that when each opportunity is presented the agent can perfectly perceive r_{is} and has no risk of being stopped or captured while committing the theft, only after consuming the value r_{is}.

His decision to steal or not is a very simple cost/benefit analysis of the two paths.

For the first opportunity, $s = 1$, if he does not take opportunity r_{is}, his expected payoff is simply the future stream of revenue from human capital investment and "rental":

$$\sum_{t=2}^{T} E(\beta_i^t(1-\bar{x}_{it})K_{it})$$

For any subsequent opportunity, the payoff is the future stream of payments, net of any possibility loss due to human capital "tarnishing" from previous crimes,

$$(1-P_i)\left(\sum_{t=2}^{T} E(\beta_i^t(1-\bar{x}_{it})K_{it})\right) + P_i\left(\sum_{t=2}^{T} E(\beta_i^t(1-\bar{x}_{(\tau)it})K_{(\tau)it})\right)$$

where the probability of punishment is $P_i = (1-(1-p_{i\bar{s}})^{\bar{s}})$, with $\bar{s}$ being the number of acts of theft undertaken to that point in the sequence, the damage to his human capital in case of punishment is τ and $\{\bar{x}_{it}\}_2^T$ is the sequence of human capital investment choices that maximizes his payoff contingent on any future shocks to human capital levels.

If he does take the opportunity r_{is}, his expected payoff is

$$r_{is}+(1-P_i)(1-p_{is})\left(\sum_{t=2}^{T} E(\beta_i^t(1-\bar{x}_{it})K_{it})\right)+(1-(1-P_i)(1-p_{is}))\left(\sum_{t=2}^{T} E(\beta_i^t(1-\bar{x}_{(\tau)it})K_{(\tau)it})\right)+\beta_i^{\delta}p_{is}c_{it}+$$

where

- p_{is} is the probability of being caught and punished by the criminal justice system. For simplicity, I do not try to separate out the impact of being arrested vs being charged, being charged vs being convicted, being convicted vs being sentenced, etc.

- $c_{it} < 0$ is his disutility from any detection, capture or punishment (including any social stigma from an arrest record, a conviction record, etc).

- $\gamma_i(1 - x_{i1})K_{i1}, \gamma_i < 0$ is the opportunity cost of the crime, γ_i is the amount of time it takes

- (τ) marks the shift in his human capital K_{it}, his human capital production parameters d_{0i}, d_{1i}, and his human capital investment strategy x_{it} if he is caught and develops a criminal justice record that limits employers willingness to hire and pay him. Thus, $\tau \in (0, 1)$ and $K_{(\tau)it} = (1 - \tau)K_{it}$

- δ is a period, less than a year, before punishment, so $\delta \in (0, 1)$.

A few other modeling decisions are worth brief comments:

Theft opportunities are modeled as random processes, where the distribution is the same across all individuals. The logic behind this is developed in another paper, but a brief outline of the logic is that per-act earnings from theft in the NLSY 1997 follow a roughly lognormal distribution (see figure 4.1). Average earnings per theft are low, and negatively correlated with number of acts (i.e., people who steal frequently tend to make less than people who steal rarely). There does not seem to be any stability in per-act earnings from year to year, and correlations in earnings from theft are low between siblings. While it is difficult to prove a negative, it does not appear to be any “theft ability”.

One could potentially develop a process by which expected risk of punishment is updated over time, or by which an initial tarnishing (from a criminal record) can be worsened by a second offense. These are interesting future extensions of the model, and may be helpful in looking at other types of crime. However, theft behavior in the NLSY is very short-lived, with most thieves only active in a single year, and so it is difficult to estimate a more complicated dynamic model.

4.3.3 Sequence of Actions

Actions in the first period 1 can thus be divided into two stages. In the first stage, if there has been a shock to his human capital level in the previous period, the agent recomputes the policy function for optimal investment, sets x_{it} accordingly and invests. In the second stage, the agent is exposed to a sequence of s opportunities, and takes any which offer a return greater than the expected cost.

All other periods, $t \in \{2, \ldots, T\}$, the agent simply sets his human capital investment level.

4.3.4 Solving the Model

Optimal Human Capital Investment

The human capital optimization problem is a dynamic optimization problem, and the policy function for the optimal x_{it} is

$$x_{i\hat{t}} = \frac{1}{K_{i\hat{t}}} \left(d_{0i} d_{1i} \beta_i \frac{1 - \beta_i^{T-\hat{t}}}{1 - \beta_i} \right)^{\frac{1}{(1-d_{1i})}}$$

for $\hat{t} \neq T$ (if $\hat{t} = T$, then $x_{i\hat{t}} = 0$).

As would be expected, optimal human capital investment for agent i in period $\hat{t}$ is increasing in time preference β and productivity of investment (d_0, d_1) and decreasing in $T - \hat{t}$. Perhaps a bit surprising, it is decreasing in $K_{i\hat{t}}$. The intuition for this is actually straightforward: in this case human capital is specifically marketable skills, as opposed to general potential. It is a measure of the money agent i can make right now, and so higher values of K_{it} are direct measures of opportunity cost of investment. The production parameters d_0, d_1 measure the sensitivity of future growth in marketable skills to the level of investment, and are thus more closely tied to the idea of human capital as general potential.

Optimal Response to Theft Opportunities

It is advantageous to put r_{is} on one side, as the direct benefit of stealing, and put all the other terms on the other side, simplifying a bit, to measure the cost of stealing. We then get an indicator function c that becomes the optimal policy in response to opportunity r_{is} at point s in period t.

$$c = \begin{cases} 1 \quad \text{if} & r_{is} \geq \frac{p_{is}}{1-P_i} \sum_{t=2}^{T} \beta_i^t \left((1-\bar{x}_{it})K_{it} - (1-\bar{x}_{(\tau)it})K_{(\tau)it} \right) \\ & \quad +\beta_i^{\delta} p_{is} c_i + \gamma_i (1 - x_{i\hat{t}}) K_{i1} \\ 0 \quad \text{else} & \end{cases}$$

As constructed, the optimal choice in response to every criminal opportunity r_{ist} is to "play" c; that is, to steal if and only if the return is greater than the future cost. The probability of stealing is decreasing in all variables that drive the right hand side costs: $p, \gamma, c_{it}, \tau_{it}, K_{it}, \beta_i, d_{0i}, d_{1i}$.

Optimal Strategy

The optimal strategy facing any criminal opportunity r_{is}, thus becomes

$$\{\{\bar{x}_{it}\}_2^T, c(r_{is})\}$$

Similar to the opportunity-specific indicator function $c(r_{is})$, we can define an indicator function $\bar{e}(\max r_{is})$ for the entire sequence of opportunities, such that

$$\bar{e} = \begin{cases} 1 \quad \text{if} & \max r_{is} \geq p_{is} \sum_{t=2}^{T} \beta_i^t \left((1-\bar{x}_{it})K_{it} - (1-\bar{x}_{(\tau)it})K_{(\tau)it} \right) + p_{is} c_i + \gamma_i (1 - x_{i\hat{t}}) K_{i\hat{t}} \\ 0 \quad \text{else} & \end{cases}$$

4.3.5 Predictions of the Model Under Different Parameter Values

As discussed above, theft will take place if and only if the value of an opportunity is greater than or equal to the expected cost of taking it. Given a baseline parameterization where theft does not occur, the question becomes what parameterizations would lead to a systematically higher probability of theft occurring?

The obvious possibilities are:

1. Lower cost of punishment c_i
2. Lower expected probability of punishment p
3. Lower time to exploit a theft opportunity γ_i
4. Lower value of initial capital, K_0
5. Lower human capital production potential via lower values of d_0, d_1
6. Lower time discounting β
7. Lower damage to human capital τ

All of these parameters are very difficult to measure objectively in even detailed panel data sets. However, we can make plausible inferences in some cases. I review the different possibilities.

The data set used in this paper, the NLSY 1997, asks respondents a number of questions about predicted probability of punishment if they were to commit a crime (a rough analog of p) , but has no data on the disutility they would experience from going to jail c_i or the time it takes them to steal (or would take them to steal if they did) γ_i. My ability to compare c_i and γ_i values across respondents is therefore limited.

As discussed in the literature review, the damage to human capital τ from a conviction record is hard to measure. In theory it may be variable across populations and so may be a difference in parameterization. However, it seems unlikely to vary substantially. I do not examination variation in this parameter as a possible difference between thieves and non-thieves.

The next three parameters, initial capital K_0, discount rate β and human capital production parameters d_0 and d_1, while effectively impossible to measure directly, will induce certain changes in measurable variables such as wages, hours worked and years of education. In table 4.1 and in figures 4.2, 4.3 and 4.4 I show the development of human capital investment, human capital, and wages under an illustrative range of parameterizations.

Three things are worth noting about the different parametrizations.

The first is the development of wages over the agent's "lifetime" (i.e. to period T). Figures 4.2, 4.3 and 4.4 give a straightforward sense of the overall pattern. Low initial human capital (K_0) acts to reduce the overall development of human capital K_t and wages $(1 - x_t)K_t$ while low discount rates and low human capital production both create a "crossing" pattern, where wages start high relative to the baseline but then develop more slowly and drop below the baseline wages. The intuition here is simple: in the case of low K_0, investment levels are higher, and there is some catchup effect, but it is not worth the loss of present wages to completely catchup to the baseline. In the case of low discount rate β or low human capital production parameters d_0 and d_1, the utility payoff of human capital investment is lower than in the baseline, and so human capital investment is lower over all periods. This has the effect of increasing wages in the early periods but reducing them in later periods.

As discussed below, the actual development of thieves' wages is most consistent with the low β development, in that it starts above the non-thieves in the mid-teen years but then falls in relation to non-thieves.

The second is the impact of the parameterizations on the loss of utility from a criminal record. While there is no definite consensus of the impact of conviction on wages or employment, the highest estimates are roughly a 30% drop in employment prospects, and most are much more conservative (see the literature review for a further discussion). I compute what impact a loss of 10% of human capital between periods 0 and 1 would have on expected discounted period 0 earnings. I then compare this impact across the four parameterizations. The loss of human capital has the least period 0 impact in the case of the low β parametrization, where it is roughly 1/4 what would happen in the baseline and 1/2 what would happen in a low K_0 parametrization.

This is further evidence that a low β parametrization is the best explanation of theft behavior under the assumptions of the model. Overall cost of committing theft is more elastic with regard to β than with regard to other parameters. Thus a lower cost of theft is most easily explained in the data by lower discount rates.

The third and final note is that while the exact development changes as the parameters are changed, the development of the low K_0, low β and low coefficent parametrizations relative to the baseline is very similar across a wide range of parametrizations. That is to say, I have not "cherry-picked" my examples but chose one of many ones that would make the same point.

4.4 Empirical Analysis

In this section I begin with a general discussion of the data, including summary statistics comparing thieves and non-thieves. I then present linear and semiparametric kernel regression output that shows a consistent underinvestment in human capital by thieves. I look at how predicted odds of arrest in the NLSY 1997 is related with (a) the probability of an individual stealing and (b) the probability of them being caught if they steal.

4.4.1 Overview of Data

The NLSY 1997 cohort tracks 8,984 individuals from 1997 onwards. The majority of respondents, roughly 3/4, come from a simple random sample of the United States youth population in 1997. The remaining one quarter of respondents come from an over-sampling of ethnic minority populations, specifically black and non-black hispanics. The respondents were equally ranged from the age of 12 to 16 in the initial round in 1997. The rate of attrition across years is quite low both for thieves and non-thieves.

In addition to a wide range of questions about work, earnings, education, assets, beliefs, health, family and other issues, every round of the NLSY 1997 includes a self-administered questionaire that asks respondents about potentially compromising issues such as criminal behavior. There are two sets of questions on theft in the 1997-2003 rounds: respondents are asked a number of questions about acts of theft where they stole items worth less than \$50, and a second set of questions about acts of theft where they stole items worth more than \$50. For each they are asked what methods they used to steal and (in most rounds) how many times they stole in the past year. For their theft of objects worth more than \$50 they are then asked for an estimate of the total earnings: "In [year], what was the amount of cash you received for the items you stole or would have received if you had sold them?". While many respondents admit to both types of theft, substantially more respondents, 3,684 or nearly one-half, admit to stealing at some point. Only 1,202, or roughly 12.5% of the respondents, admit to stealing items worth more than \$50 at any point. There is not a perfect division between the two groups - many of the respondents who admit to stealing low-value items say they used weapons or snatched property from their victims, which in many states would make the theft a felony, and many respondents report average earnings per theft of \$0 or nearly \$0 for their theft of items worth more than \$50. In this paper I focus on the data for the thefts worth more than \$50. Unless explicitly stated, by "thief" I am referring to a respondent who admitted to theft of at least one item worth more than \$50, and by "theft" I am referring an act which the respondent characterized as of an item(s) worth more than \$50 (which in many cases leads to a take of significantly less than \$50).

4.4.2 Comparison of Thieves and Non-Thieves

In table 4.2 I compare some basic attributes of thieves and non-thieves in the data. There is some evidence that thieves have lower potential human capital - they come from slightly poorer homes, on average, and have lower evaluated test scores and mental health. The differences are significant but not striking.

Looking at average wages and weeks worked between thieves and non-thieves in figures 4.7 and 4.8, we can see slight differences but there is not a consistent deficit in wages and/or weeks worked as would be predicted by a low K_0 parametrization. Instead, wages start above non-thieves and then drop relative to non-thieves. Weeks worked follow a similar pattern.

This is extremely different from other groupings that are possible in the data. For example in figures 4.5 and 4.6 we can see a very different pattern, and one that seems very similar to the model of low K_0 in figure 4.2. Individuals from households with annual incomes below \$50,000 show consistently lower wages and weeks worked across their teens and twenties than individuals from households with annual incomes above \$50,000. Similar figures can be drawn based on other groupings that are correlated with or indicative of differences in socio-economic status, for example grouping by race.

4.4.3 Other Evidence on Human Capital Investment Differences Between Thieves and Non-Thieves

Using a range of direct measure of human capital investment, I further substantiate that under-investment by thieves plays a significant role.

Looking at enrollment levels by year in 4.9, we can see that enrollment for thieves drops both absolutely and relative to non-thieves at the end of high school. Note that both groups are individuals without criminal convictions, eliminating incarceration effects (which tend to be minor in the NLSY generally).

Even more striking, thieves have much higher likelihood of repeating one or more grade, even holding constant for a range of other measures. I run a series of regressions where the dependent variable is number of grades skipped as of 2003. The results are given in Table 4.3. A number of variables are highly significant including test scores, receiving good grades in eighth grade, income of family of origin and mental health. Additionally, I include gender, age in 1997 and a measure of peer misbehavior (skipping school). Even with these controls we can see that thieves are 10% more likely to skip a grade. Interestingly peer behavior has no significance. I redo the regression, removing individuals with convictions, and then individuals with charges, to demonstrate that it does not appear to be legal issues or incarceration effects that are causing the grade repeats.

Another demonstration of the present orientation is that thieves have more employers. I run a regression where the dependent variable is number of employers by age 20. The results are in Table 4.4. I include the same covariates as in the previous regression - test scores, receiving good grades in eighth grade, income of family of origin, mental health, gender, age in 1997 and a measure of peer misbehavior. There are several interesting patterns in this series of regressions. First, peer effects, not at all significant in predicting number of employers, is now strongly significant. Second, while several of the covariates that seem to be good predictors of more jobs are indicative of behavioral issues, including theft and peer misbehavior (and good grades have a negative impact), higher standardized test scores are also strongly significant, suggesting that the two main drivers of many employer are (a) behavioral issues and (b) brainpower.

Thus, we can see a strong pattern of underinvestment in schooling and at the same time evidence of a great deal of engagement in the labor market. This supports the argument that discount rates are a critical issue.

4.4.4 Regressions: Predicted Odds of Arrest

As discussed earlier, there is no direct measure of disutility from prison or from criminal justice punishment, but there are measures of expected probability of punishment. For a number of early years, respondents are asked if they stole a car, what would be the odds of being arrested? If they were arrested, what would be the odds of being let of with a fine, what would be the odds of conviction, what would be the odds doing jail time?

Interestingly, these measures are (in isolation) highly predictive of theft behavior. I run two probit regressions, where the dependent variable is whether or not an individual ever stole. The results are given in Table 4.5.

As can be seen, the response to predicted probability of arrest as of 1997 is highly significant on its own in predicting that in individual will steal (other responses to similar questions in that year and later years, while generally not as strong, usually show significance). I add a number of other variables in the second specification, variables that are generally strong predictors of theft behavior - specifically, standardized test scores, mental health as tested in 2000, and cigarette usage in ages 16-17, as well as two controls, household income and age in 1997. In the second specification, the respondent's predicted odds of arrest goes from being highly significant to marginally insignificant. It seems to be a strong predictor of the decision to steal.

However, does predicted odds of arrest help to predict an individual's probability of being arrested upon committing a crime? To test this, I run another two probit regressions where the dependent variable is whether or not an individual is ever arrested. The results of the two specifications are in 4.6. In the first specification I include number of acts of theft from 1997 to 2003 and the square of the number of acts. These are both highly significant, but predicted odds of arrest is not. It remains insignificant in the second specification, where I add the same predictors of theft behavior and controls as in the previous paragraph.

It appears that the subjective measure p is strongly biased by the same impatience and present orientation that goes into β. It is highly correlated with the decision to steal, but does not appear to be correlated with objective probability.

4.4.5 Opportunity Cost

The fact that thieves appear to have identical or higher wages than non-thieves suggests that the opportunity cost component of the "cost" function is minor and plays no significant role.

4.5 Conclusion

This paper proposes a model that integrates the decision to steal with a larger question of human capital formation. There are several parameters that could potentially explain why some people steal and others don't:

1. Lower cost of punishment c_i
2. Lower expected probability of punishment p
3. Lower time to exploit a theft opportunity γ_i
4. Lower value of initial capital, K_0
5. Lower human capital production potential via lower values of d_0, d_1
6. Lower time discounting β
7. Lower damage to human capital τ

Several we have direct or indirect measures of, others we have no access to.

There are no good proxies for $c_i, \gamma_i, d_0, d_1, \tau$, so we cannot measure them. Damage to human capital τ and lost time stealing γ_i seem unlikely to vary greatly.

We have a measure of expected probability of punishment p, but this only seems to measure confidence, not objective reality.

Differences in K_0 between thieves and non-thieves do not seem to be important in the teen years, in contrast to other ways of dividing the NLSY respondents, such as by socio-economic background or ethnicity, where there are statistically significant differences in wages and hours worked strongly consistent with differences in K_0.

There are a number of patterns that strongly supports the theory that differences in β are an important part of the decision to steal. First, the development in wages and hours of thieves relative to non-thieves follows the pattern predicted by the Ben-Porath model. Secondly, differences in β lead to much greater differences in the net present value of future human capital losses than comparable differences in K_0, d_0 or d_1. Thirdly, thieves show much less investment in education or job tenure than non-thieves, which is strongly consistent with underinvestment in human capital and a lower β.

In all, the paper supports the theory that impatience is a key causal factor in the decision to steal.

Wage Development In A Range of Illustrative Parametrizations

Baseline Model

Parameter		Variable	1	2	3	4	5
					Periods		
β	0.8	Investment level x_t	0.03	0.02	0.01	0.00	0.00
d_0	0.5	Human Capital K_t	10.00	10.30	10.54	10.73	10.83
d_1	0.5	Wage $(1-x_t)K_t$	9.65	10.05	10.41	10.68	10.83
T	5	Ln(Wage) $ln((1-x-t)K_t)$	2.27	2.31	2.34	2.37	2.38
K_0	10						
Lost income from a 10% drop in human capital			0.77	0.63	0.52	0.42	0.34
						Total	2.690

Low Initial Human Capital K_0 Model

Parameter		Variable	1	2	3	4	5
					Periods		
β	0.8	Investment level x_t	0.07	0.05	0.03	0.01	0.00
d_0	0.5	Human Capital K_t	5.00	5.30	5.55	5.74	5.84
d_1	0.5	Wage $(1-x_t)K_t$	4.65	5.04	5.40	5.69	5.84
T	5	Ln(Wage) $ln((1-x-t)K_t)$	1.54	1.62	1.69	1.74	1.77
K_0	5						
Lost income from a 10% drop in human capital			0.37	0.31	0.26	0.22	0.18
						Total	1.35
						% of Baseline	50.0%

Low Discount Rate β Model

Parameter		Variable	1	2	3	4	5
					Periods		
β	0.4	Investment level x_t	0.00	0.00	0.00	0.00	0.00
d_0	0.5	Human Capital K_t	10.00	10.08	10.16	10.23	10.28
d_1	0.5	Wage $(1-x_t)K_t$	9.97	10.06	10.14	10.22	10.28
T	5	Ln(Wage) $ln((1-x-t)K_t)$	2.30	2.31	2.32	2.32	2.33
K_0	10						
Lost income from a 10% drop in human capital			0.40	0.16	0.06	0.03	0.01
						Total	0.66
						% of Baseline	24.5%

Low Human Capital Production Coefficients Model

Parameter		Variable	1	2	3	4	5
					Periods		
β	0.8	Investment level x_t	0.01	0.01	0.00	0.00	0.00
d_0	0.25	Human Capital K_t	10.00	10.13	10.26	10.37	10.46
d_1	0.25	Wage $(1-x_t)K_t$	9.92	10.07	10.22	10.35	10.46
T	5	Ln(Wage) $ln((1-x-t)K_t)$	2.29	2.31	2.32	2.34	2.35
K_0	10						
Lost income from a 10% drop in human capital			0.79	0.64	0.51	0.41	0.33
						Total	2.69
						% of Baseline	100.0%

Table 4.1: Optimal human capital investment, development and wages under four parameterizations of the model, for five periods. Additionally, for each parametrization the discounted lost income in period 0 from a 10% drop in K (due to a conviction or criminal record before period 1 begins) is computed. Note that in the low K_0 parametrization the lost value of lifetime income in case of a criminal record before period 0 is roughly 50% of that under the baseline, in the low β parametrization the lost value is less than 25% of that under the baseline, and in the low coefficients parametrization it is roughly the same as in the baseline. In other words, lower β is most consistent with a “thief” parameterization, where the costs of theft are lower than the norm.

Summary Statistics			
	All	Non-Thieves	Thieves
Age in 1997	14.85	14.86	14.80
(St. Dev)	(1.45)	(1.45)	(1.42)
Household of origin income	46,361	47,086	41,702
(St. Dev)	(42,143)	(42,656)	(38,380)
White	51.9%	52.2%	50.1%
Black	26.0%	25.9%	26.6%
Hispanic	21.1%	21.0%	22.4%
Female	48.8%	51.4%	32.0%
ASVAB Scores (Math and Verbal)	45,445	46,555	38,100
(St. Dev)	(29,115)	(29,209)	(27,388)
Mental Health (20 best, 0 worst)	15.32	15.40	14.82
(St. Dev)	(2.55)	(2.50)	(2.79)

Table 4.2: A comparison of basic data between thieves and non-thieves. Thieves are more likely to be male, come of a family of origin with lower income and have lower test scores and evaluated mental health. No age or ethnicity differences emerge in the data.

Linear Regression
Dependent Variable: Number of grades repeated by 2003

		Eliminating respondents with convictions	Eliminating respondents with charges
Thief	0.086***	0.078**	0.094**
	(3.514)	(2.718)	(3.002)
Age in 1997	0.014*	0.012	0.009
	(2.317)	(1.857)	(1.460)
Household Income	-0.001***	-0.001***	-0.001**
($1000s)	(-3.881)	(-3.578)	(-3.213)
ASVAB Math/Verbal	-0.005***	-0.005***	-0.005***
(1000 pts)	(-15.276)	(-14.512)	(-13.769)
Good grades (8th grade)	-0.115***	-0.107***	-0.109***
	(-5.903)	(-5.275)	(-5.238)
Female	-0.055***	-0.041*	-0.032
	(-3.314)	(-2.425)	(-1.840)
Peers cut classes in 1997	0.005	0.001	0.001
	(0.740)	(0.106)	(0.104)
Constant	0.387***	0.408***	0.422***
	(4.420)	(4.570)	(4.637)
R-squared	0.177	0.166	0.161
N	2955	2608	2449

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
t-statistics in parentheses

Table 4.3: Regression of number of grades repeated in 2003. As can be seen, even controlling for a range of measures of academic preparation and class, thieves are more likely to repeat grades.

Linear Regression
Dependent Variable: Number of jobs by age 20

		Eliminating respondents with charges	Eliminating respondents with convictions
	est1	est3	est2
Thief	0.818***	0.953***	0.887***
	(7.093)	(6.849)	(5.832)
Age in 1997	0.094**	0.090**	0.088**
	(3.200)	(2.947)	(2.812)
Household Income	-0.001	-0.000	-0.000
($1000s)	(-0.575)	(-0.075)	(-0.016)
ASVAB Math/Verbal	0.010***	0.009***	0.009***
(1000 pts)	(6.171)	(5.122)	(5.158)
Good grades (8th grade)	-0.317***	-0.287**	-0.264**
	(-3.424)	(-2.930)	(-2.626)
Female	0.041	0.100	0.070
	(0.515)	(1.204)	(0.820)
Peers cut classes in 1997	0.222***	0.224***	0.233***
	(6.419)	(6.133)	(6.216)
Constant	1.861***	1.813***	1.772***
	(4.472)	(4.175)	(3.990)
R-squared	0.042	0.041	0.039
N	3704	3203	2994

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
t-statistics in parentheses

Table 4.4: Regression of number of jobs by age 20. As can be seen, even controlling for a range of measures of academic preparation and class, thieves have nearly one more job than non-thieves.

Probit Estimation
Dependent variable: Committed theft ever, 1997-2003

	Arrest Prediction Only		Full Probit Model		
	Coefficients	Marginal Effect	Coefficients	Marginal Effect	$\bar{X}$
Rs predicted odds of arrest	-0.003**	-0.001**	-0.002	-0.000	64.401
1997	(-2.64)	(-2.65)	(-1.95)	(-1.95)	
ASVAB Math/Verbal			-0.006***	-0.001***	57.049
(1000 pts)			(-3.59)	(-3.62)	
Household Income			-0.000	-0.000	61.003
($1000s)			(-0.37)	(-0.37)	
Mental Health			-0.031	-0.007	15.769
(eval 2000)			(-1.78)	(-1.78)	
Days smoking in			0.027***	0.006***	7.240
last month (age 16/17)			(8.50)	(8.49)	
Age in 1997			0.068*	0.015*	14.628
			(2.28)	(2.29)	
Constant	-0.793***		-1.279*		
	(-10.55)		(-2.38)		
N	1410	1410	1410	1410	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
t-statistics in parentheses

Table 4.5: Probit model of stealing at least once between 1997-2003 among white males. Notice that subjective predicted odds of arrest when stealing a car, in response to a question asked in 1997, is highly significant on its own and loses significance when other variables are added.

Probit Estimation
Dependent variable: Charged with theft ever, 1997-2003

	Arrest Prediction Only		Full Model		
	Coefficients	Marginal Effect	Coefficients	Marginal Effect	$\bar{X}$
Rs predicted odds of arrest	0.001	0.000	0.002	0.000	64.401
1997	(0.80)	(0.80)	(1.08)	(1.08)	
Number of acts of theft	0.064***	0.009***	0.052***	0.006***	1.528
1997-2003	(5.82)	(5.64)	(4.62)	(4.47)	
Square of number of acts of theft	-0.000***	-0.000***	-0.000**	-0.000**	75.429
	(-3.55)	(-3.52)	(-2.67)	(-2.65)	
ASVAB Math/Verbal			-0.005*	-0.001*	57.049
(1000 pts)			(-2.53)	(-2.56)	
Household Income			-0.000	-0.000	61.003
($1000s)			(-0.30)	(-0.30)	
Mental Health			-0.028	-0.003	15.769
(eval 2000)			(-1.32)	(-1.32)	
Days smoking in			0.017***	0.002***	7.240
last month (age 16/17)			(4.42)	(4.45)	
Age in 1997			0.043	0.005	14.628
			(1.17)	(1.17)	
Constant	-1.593***		-1.695*		
	(-15.09)		(-2.54)		
N	1410	1410	1410	1410	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
t-statistics in parentheses

Table 4.6: Probit model of ever being charged with stealing 1997-2003 among white males. Notice that subjective predicted odds of arrest when stealing a car, in response to a question asked in 1997, is never significant as a predictor of actually being arrested.

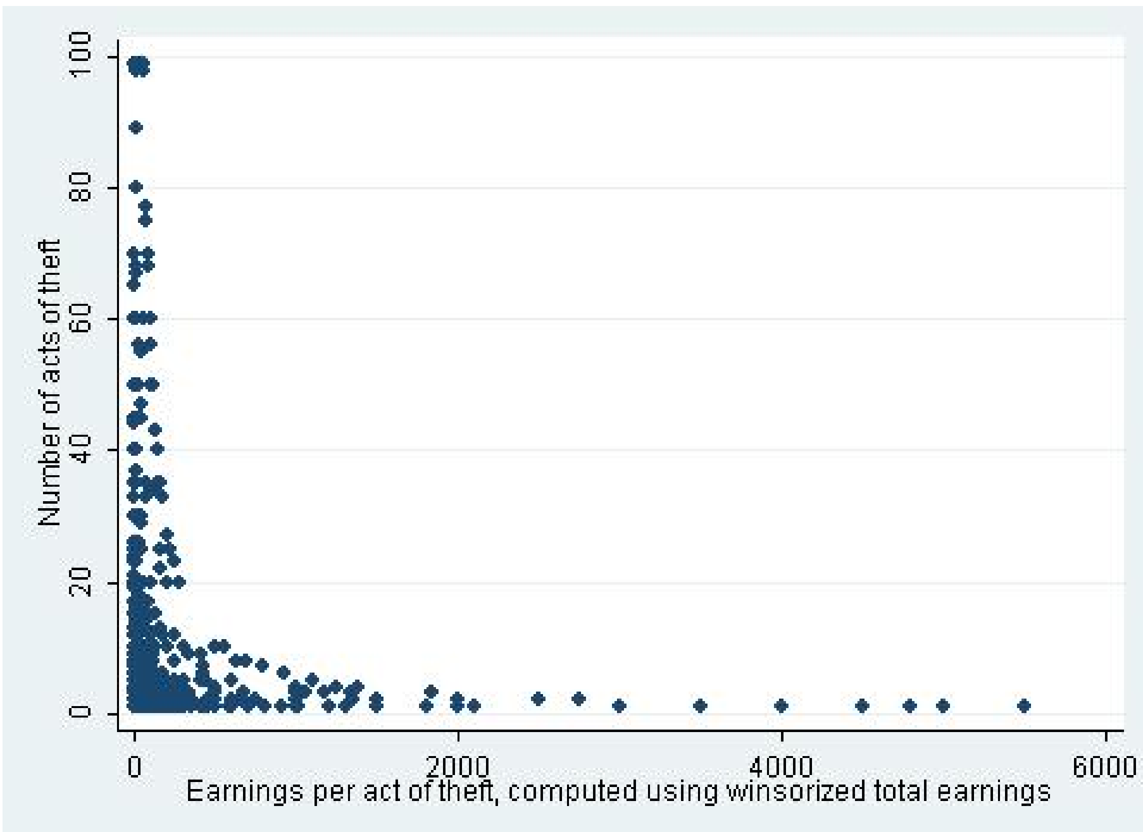

Figure 4.1: Data on earnings from theft and number of acts in the NLSY, after winsorizing average earnings per theft at the 95% level.

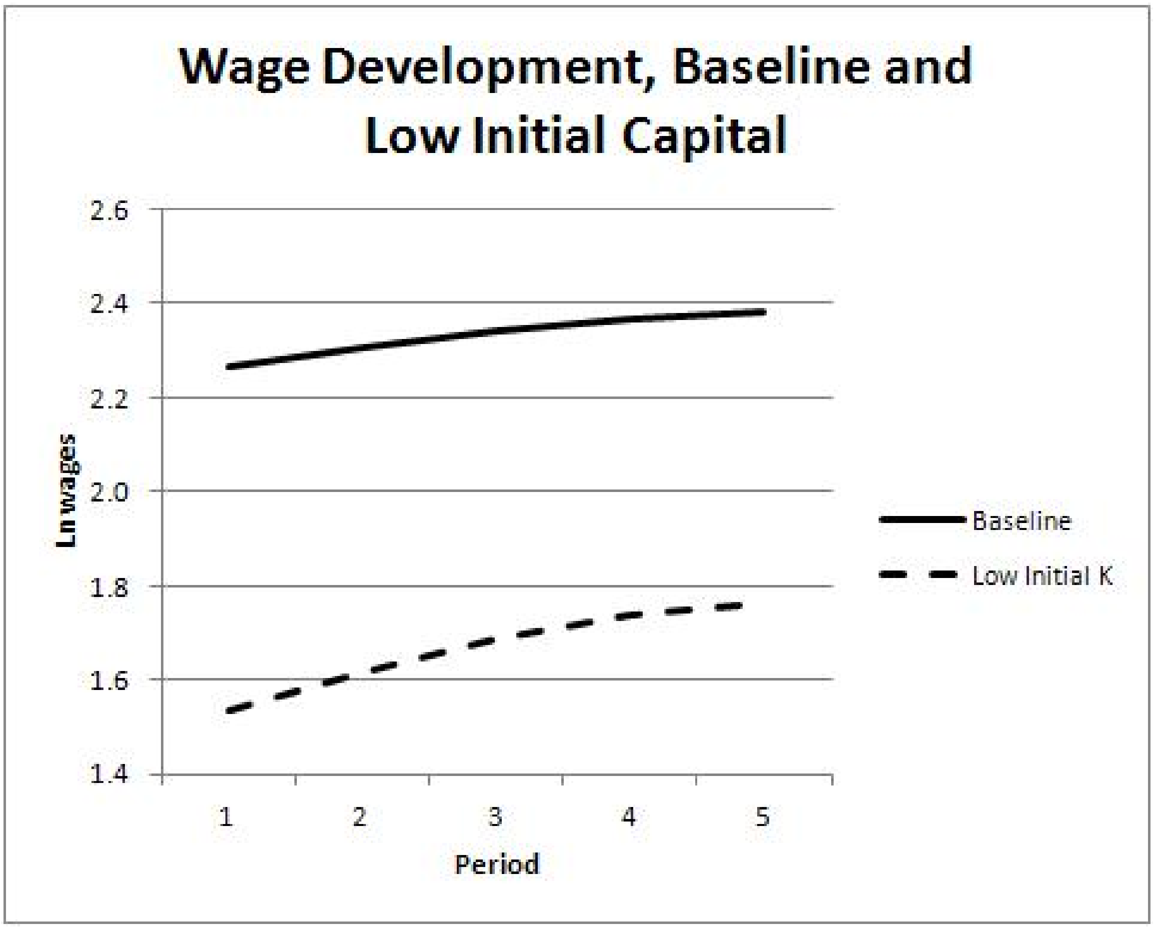

Figure 4.2: Wage development in the human capital model outlined in the text. The baseline parameterization is compared with the low K_0 parameterization (where K_0 is 1/2 the level in the baseline). As can be seen, wages are roughly halved in the low K_0 version.

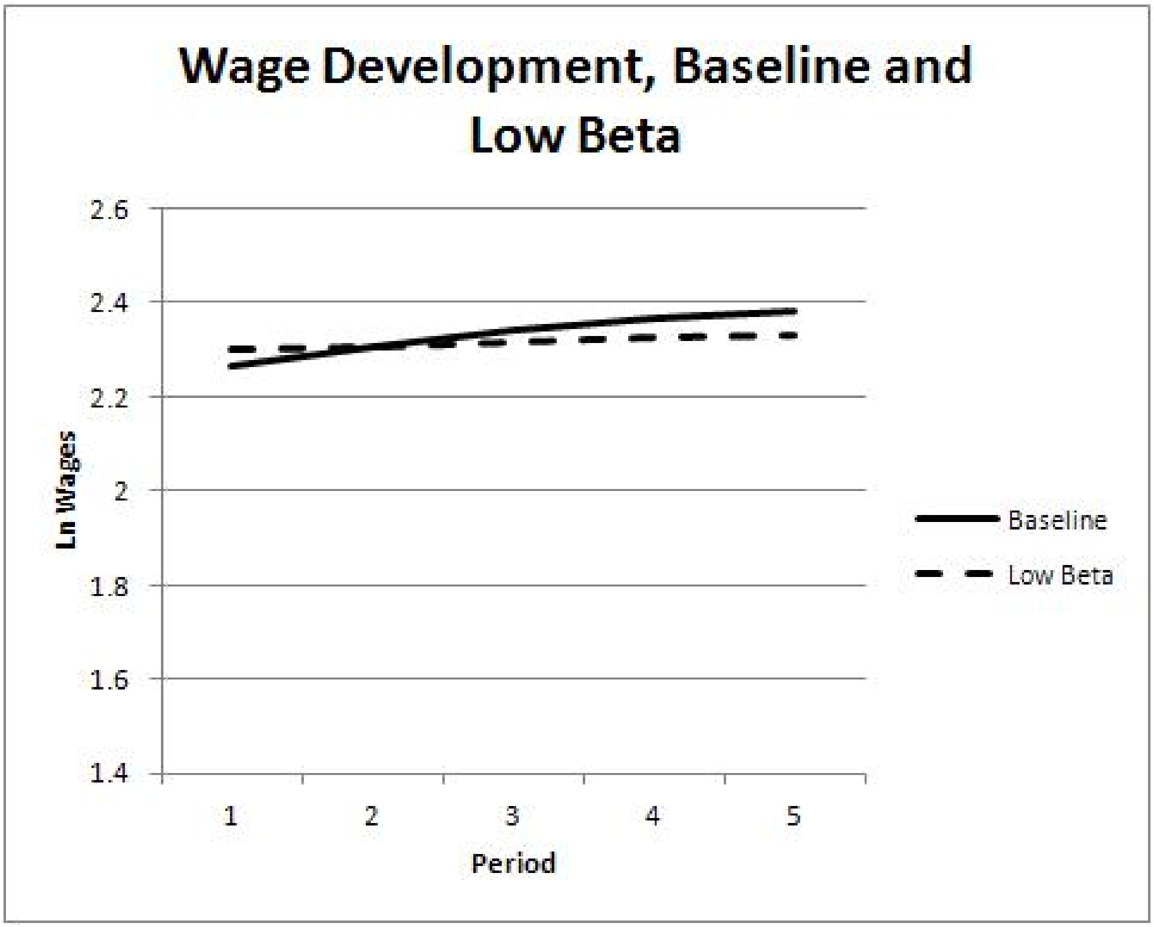

Figure 4.3: Wage development in the human capital model outlined in the text. The baseline parameterization is compared with the low β parameterization (where β is 1/2 the level in the baseline). As can be seen, wages are extremely similar, but start slightly higher in the low β version, dropping relative to the baseline. This is because human capital investment yields a lower discounted value for a low β individual.

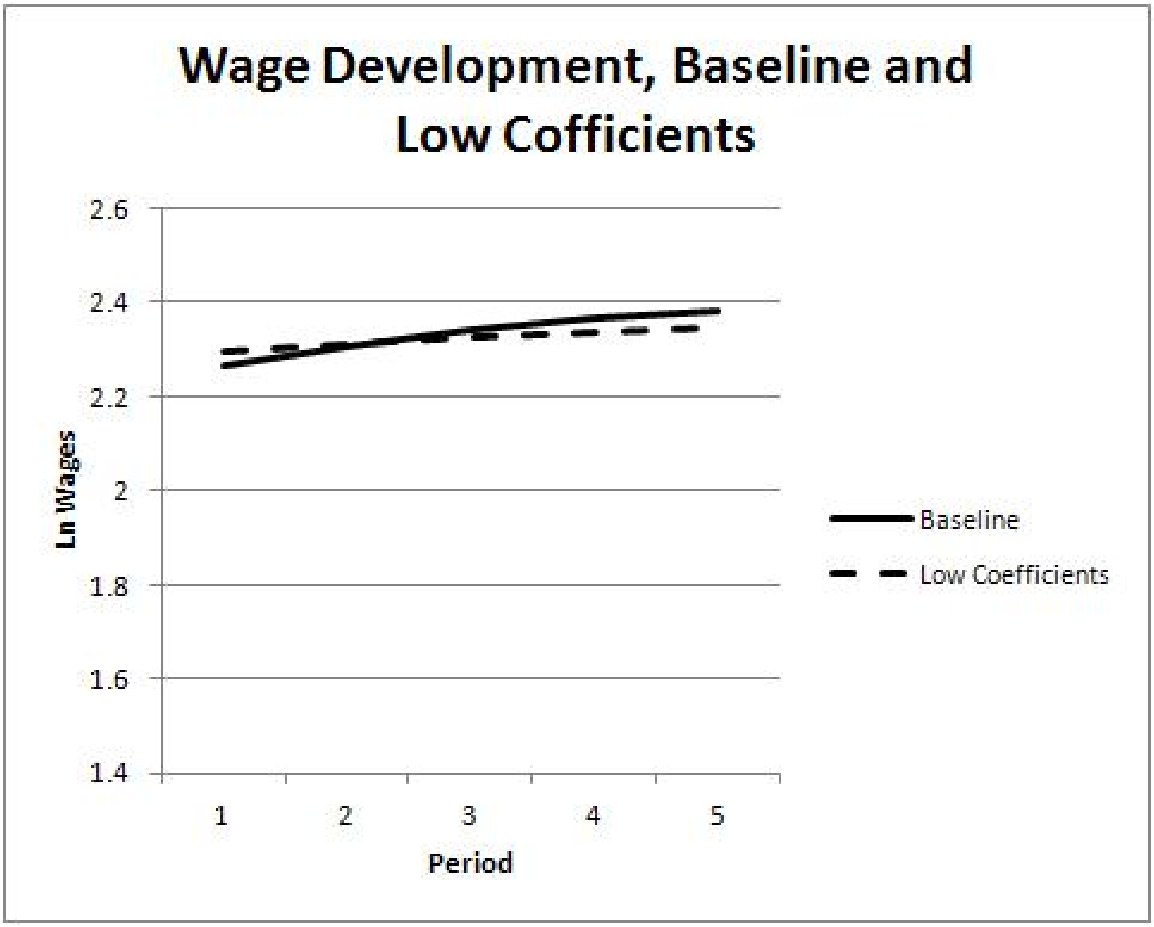

Figure 4.4: Wage development in the human capital model outlined in the text. The baseline parameterization is compared with the low d_0 and d_1 parameterization (where d_0 and d_1 are 1/2 the level in the baseline). As can be seen, wages are extremely similar, but start slightly higher in the low d_0 and d_1 version, dropping relative to the baseline. This is because human capital investment yields a lower discounted value for an individual with lower d_0 and d_1 values.

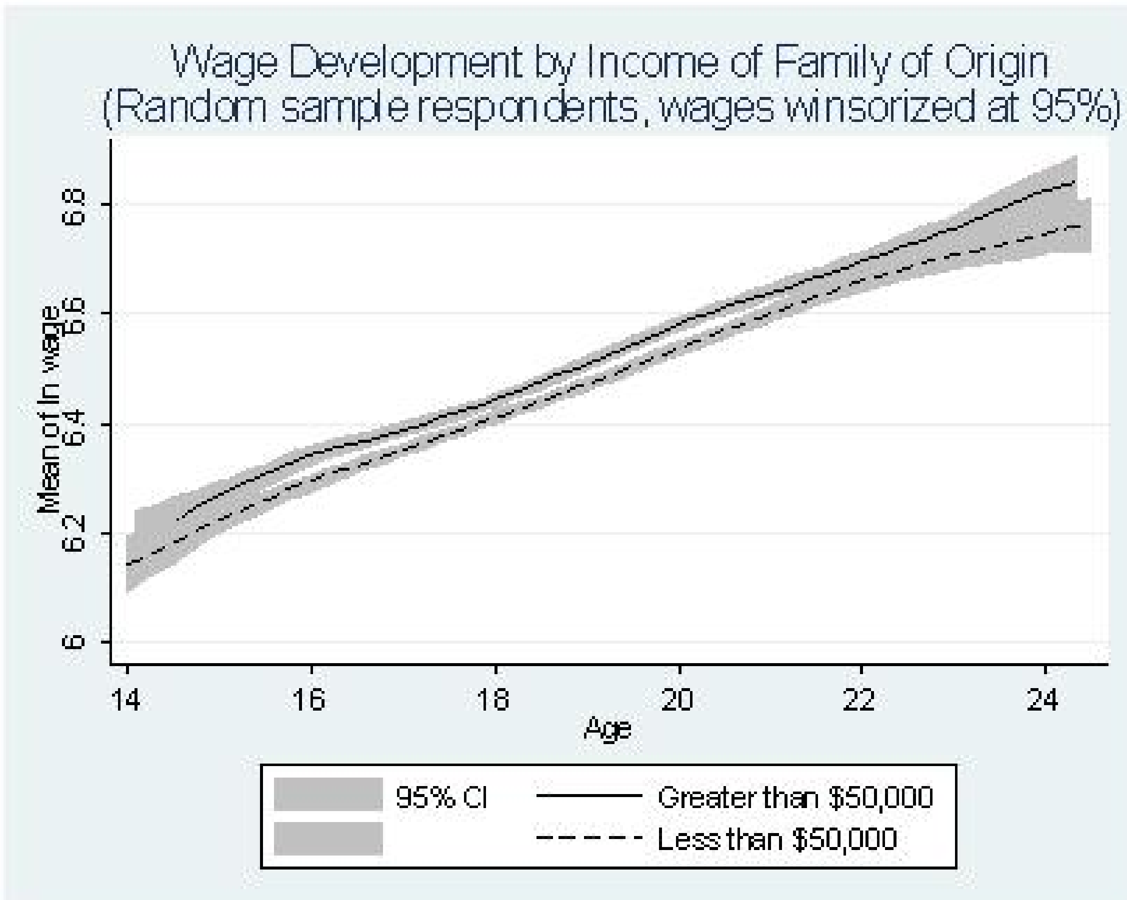

Figure 4.5: Actual wage development in the NLSY 1997 compared for respondents divided by income of household of origin. Notice that wages are solidly lower for lower income respondents throughout the teens and early twenties. The “signature” is very similar to that of the low K_0 figure. Similar figures can be drawn for a range of groupings highly correlated with socioeconomic status, for example comparing across races.

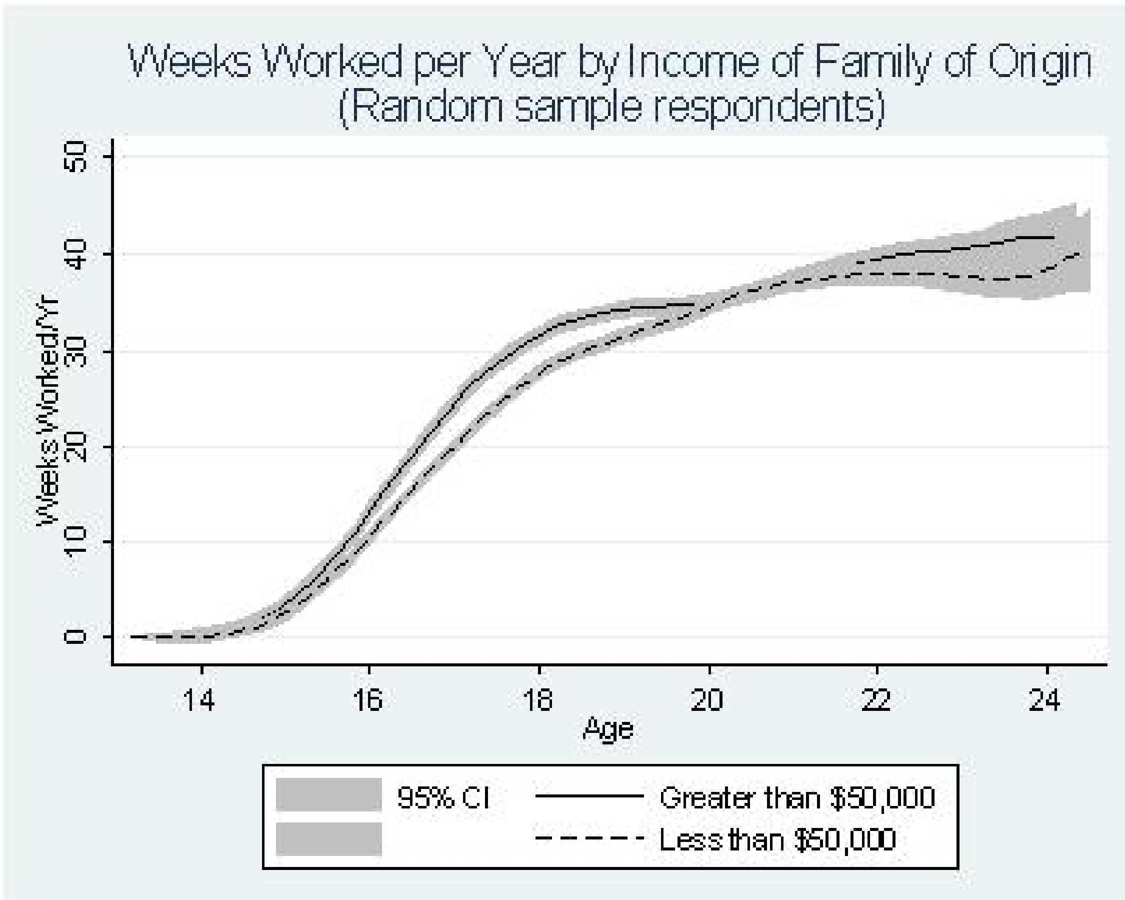

Figure 4.6: Actual development of weeks worked per year in the NLSY 1997 compared for respondents divided by income of household of origin. Notice that weeks worked are lower for lower income respondents throughout the teens and for some of the early twenties. Since measures of educational attainment are higher for higher income individuals, lower weeks worked appears not to be a measure of investment but of human capital. Similar figures can be drawn for a range of groupings highly correlated with socioeconomic status, for example comparing across races.

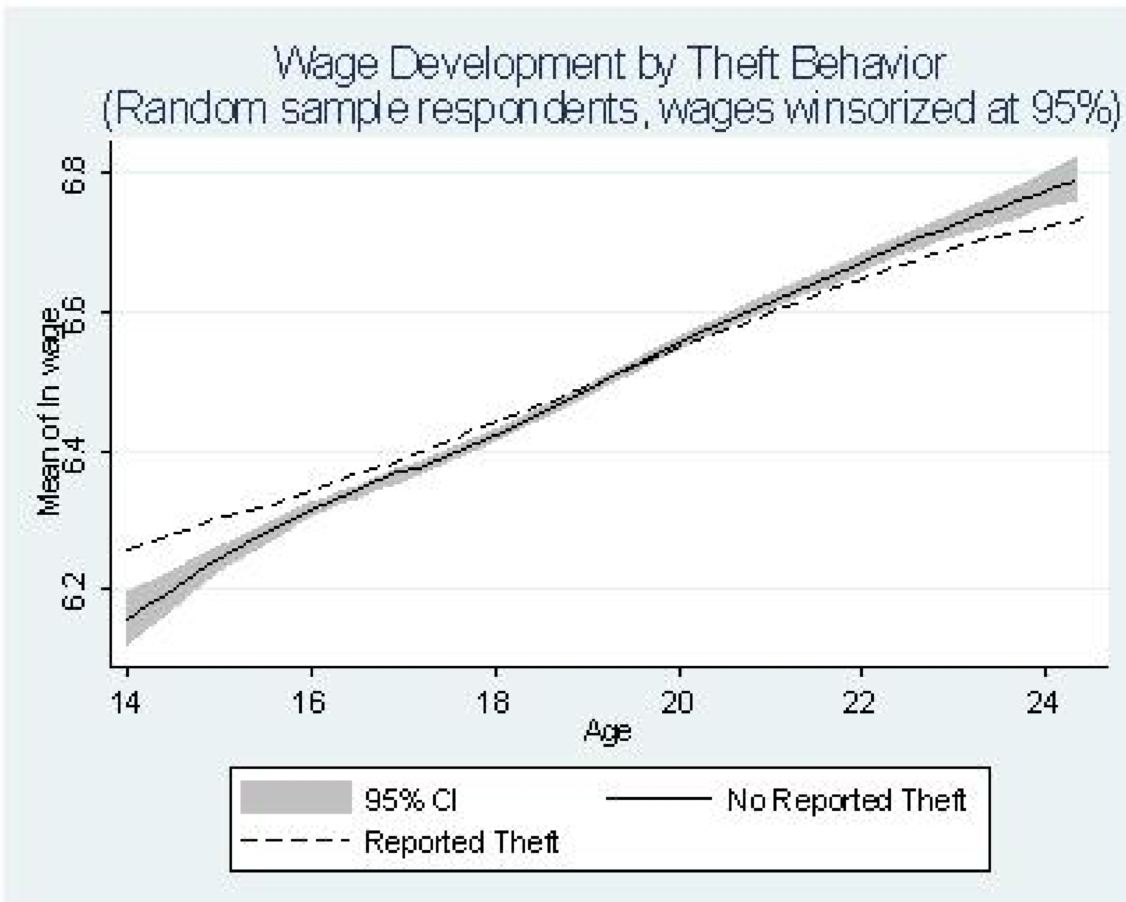

Figure 4.7: Actual wage development in the NLSY 1997 for thieves and non-thieves. Notice that wages are slightly higher for thieves in the teen years but then become lower in the twenties. The "signature" is very similar to that of the low β and low d_0 and d_1 figures.

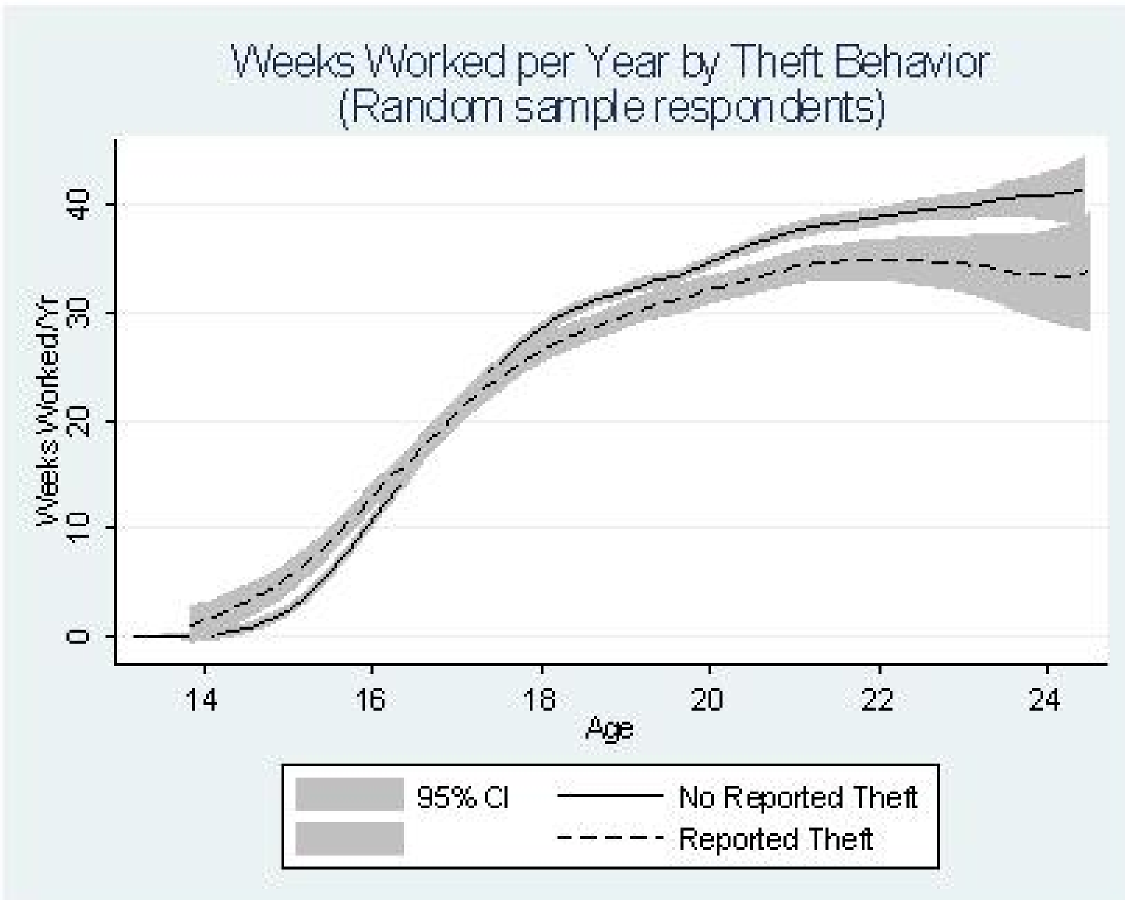

Figure 4.8: Actual development of weeks worked per year in the NLSY 1997 for thieves and non-thieves. Notice that weeks are slightly higher for thieves in the teen years but then become lower by the twenties. Since measures of educational attainment are lower for the thieves, weeks worked appears to be a measure of value to employers (i.e. human capital).

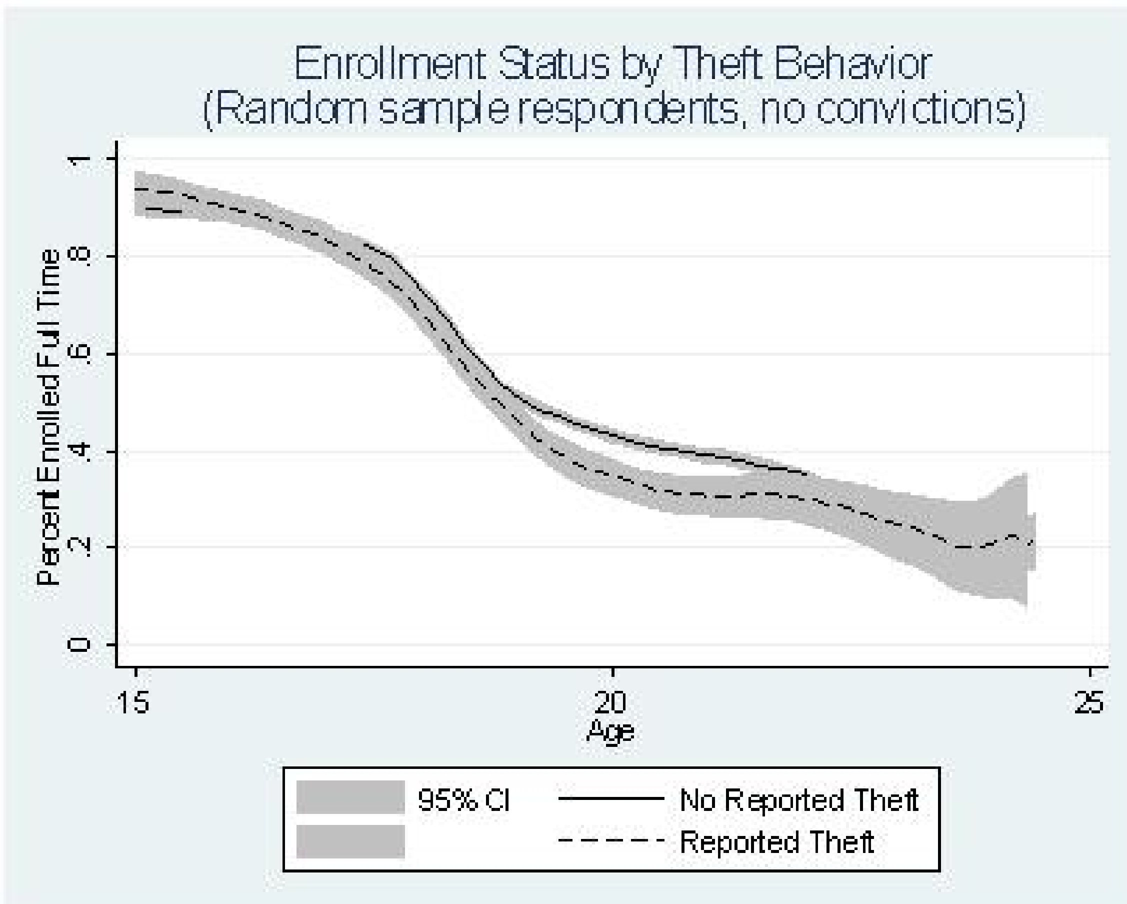

Figure 4.9: Enrollment status among thieves and non-thieves from mid-teen years to age 24, for individuals with no convictions. The enrollment of both groups drops at the end of high school, but drops more among thieves.

References

Chris Abani. The lottery, 1 February 2004 2004.

M. A. Aguiar and M. Bils. *Has Consumption Inequality Mirrored Income Inequality?*, 2011.

J. Aitchison and J. A. Brown. *The Lognormal Distribution.* Cambridge Univ. Pr., 1969.

J. D. Angrist and A. B. Krueger. Empirical strategies in labor economics. *Handbook of Labor Economics*, 3:1277–1366, 1999.

A. Barnett, A. Blumstein, and D. P. Farrington. A prospective test of a criminal career model. *Criminology*, 27(2):373–388, 1989.

G. S. Becker. Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5):9–49, 1962.

G. S. Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217, 1968.

C. R. Belfield, M. Nores, S. Barnett, and L. Schweinhart. The high/scope perry preschool program. *Journal of Human Resources*, 41(1):162–190, 2006.

Y. Ben-Porath. The production of human capital and the life cycle of earnings. *The Journal of Political Economy*, 75(4):352–365, 1967.

S. M. Berman. Limit theorems for the maximum term in stationary sequences. *The Annals of Mathematical Statistics*, 35(2):502–516, 1964.

A. Blumstein and J. Cohen. Estimation of individual crime rates from arrest records. *The Journal of Criminal Law and Criminology (1973-)*, 70(4):561–585, 1979.

R. Brame, R. Paternoster, and A. R. Piquero. Thoughts on the analysis of group-based developmental trajectories in criminology. 2011.

L. M. Broidy, D. S. Nagin, R. E. Tremblay, J. E. Bates, B. Brame, K. A. Dodge, D. Fergusson, J. L. Horwood, R. Loeber, and R. Laird. Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology*, 39(2):222–245, 2003.

A. C. Cameron and P. K. Trivedi. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1(1): 29–53, 1986.

A. C. Cameron and P. K. Trivedi. *Regression Analysis of Count Data*, volume 30. Cambridge Univ Pr, 1998.

A. C. Case and L. F. Katz. The company you keep: The effects of family and neighborhood on disadvantaged youths. *NBER Working Paper*, 1991.

Shannan Catalano. Victimization during burglary. Technical Report NCJ 227379, U.S. Department of Justice Office of Justice Programs Bureau of Justice Statistics, 2010.

E. P. Cheyney. The court of star chamber. *The American Historical Review*, 18(4):727–750, 1913.

F. Cunha and J. Heckman. The technology of skill formation. *The American Economic Review*, 97(2):31–47, 2007.

F. Cunha, J. J. Heckman, L. Lochner, and D. V. Masterov. Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812, 2006.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

A. DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer Verlag, 2008.

HA David. *Order Statistics*. Wiley, New York, first edition, 1981.

A. Deaton. Height, health, and development. *Proceedings of the National Academy of Sciences*, 104(33):13232–13237, 2007.

S. DellaVigna and M. D. Paserman. Job search and impatience. *Journal of Labor Economics*, 23(3):527–588, 2005.

Mark Doyle. 24th annual retail theft survey, 2012.

A. V. D'Unger, K. C. Land, P. L. McCall, and D. S. Nagin. How many latent classes of delinquent/criminal careers? results from mixed poisson regression analyses 1. *American Journal of Sociology*, 103(6):1593–1630, 1998.

D. P. Farrington. Age and crime. *Crime and justice*, 7:189–250, 1986.

M. Felson. *Crime and Everyday Life*. Sage Publications, Inc, 2009.

Robert M. Figlio, Paul E. Tracy, and Marvin E. Wolfgang. Delinquency in a birth cohort in philadelphia, pennsylvania, 1945-1963.

R. B. Freeman. Crime and the employment of disadvantaged youths. *NBER working paper*, 1991.

R. B. Freeman. The economics of crime. *Handbook of Labor Economics*, 5(3 PART C): 3529–3572, 1999.

M. Friedman and S. Kuznets. Income from independent professional practice. 1954.

M. Gardner and L. Steinberg. Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. *Developmental Psychology*, 41(4):625–635, 2005.

E. L. Glaeser, B. Sacerdote, and J. A. Scheinkman. Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–548, 1996.

P. W. Glimcher. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics.* The MIT Press, 2004.

M. Gottfredson and T. Hirschi. The true value of lambda would appear to be zero: An essay on career criminals, criminal careers, selective incapacitation, cohort studies, and related topics. *Criminology*, 24(2):213–234, 1986.

M. R. Gottfredson and T. Hirschi. *A General Theory of Crime.* Stanford Univ Pr, 1990.

E. D. Gould, B. A. Weinberg, and D. B. Mustard. Crime rates and local labor market opportunities in the united states: 1979-1997. *Review of Economics and Statistics*, 84(1): 45–61, 2002.

D. F. Greenberg. Modeling criminal careers. *Criminology*, 29(1):17–46, 1991.

W. Greene. *Functional Form and Heterogeneity in Models for Count Data*, volume 1. Now Pub, 2007.

William H. Greene. *Econometric Analysis.* Perason Prentice Hall, Upper Saddle River, New Jersey, 07458, 6th edition, 2008. ISBN 978-0-13-513245-6.

J. Grogger. The effect of arrests on the employment and earnings of young men. *The Quarterly Journal of Economics*, 110(1):51–71, 1995.

J. Grogger. Market wages and youth crime. *Journal of Labor Economics*, 16(4):756–791, 1998.

G. Guo, T. Cai, R. Guo, H. Wang, and K. M. Harris. The dopamine transporter gene, a spectrum of most common risky behaviors, and the legal status of the behaviors. *PLoS One*, 5(2):e9352, 2010.

J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labour market outcomes and social behaviour. *Journal of Labor Economics*, 24(3): 411–482, 2006.

T. Hirschi and M. Gottfredson. Age and the explanation of crime. *American Journal of Sociology*, 89(3):552–584, 1983.

J. R. Kling, J. Ludwig, and L. F. Katz. Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Quarterly Journal of Economics*, 120(1):87–130, 2005.

C. B. Klockars. *The Professional Fence.* Free Press New York, 1974.

J. H. Laub, D. S. Nagin, and R. J. Sampson. Trajectories of change in criminal offending: Good marriages and the desistance process. *American Sociological Review*, pages 225–238, 1998.

D. Lee and J. McCrary. Crime, punishment, and myopia. *NBER Working Paper*, 2005.

S. D. Levitt. Juvenile crime and punishment. *Journal of Political Economy*, 106(6):1156–1185, 1997.

S. D. Levitt and L. Lochner. The determinants of juvenile crime. *Risky Behavior Among Youths: An Economic Analysis*, pages 327–73, 2001.

S. D. Levitt and S. A. Venkatesh. An economic analysis of a drug-selling gang's finances. *Quarterly Journal of Economics*, 115(3):755–789, 2000.

D. Lisak and P. M. Miller. Repeat rape and multiple offending among undetected rapists. *Violence and victims*, 17(1):73–84, 2002.

L. Lochner. Education, work, and crime: A human capital approach. *International Economic Review*, 45(3):811–843, 2004.

L. Lochner and E. Moretti. The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *The American Economic Review*, 94(1):155–189, 2004.

R. B. Manning. The origins of the doctrine of sedition. *Albion: A Quarterly Journal Concerned with British Studies*, 12(2):99–121, 1980.

C. F. Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993.

S. E. Merry. Toward a general theory of gossip and scandal. *Donald Black*, 1984.

J. Mincer. Schooling, experience, and earnings. human behavior social institutions. 1974.

T. E. Moffitt. Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychology*, 100:674–701, 1993.

C. Morselli and P. Tremblay. Criminal achievement, offender networks and the benefits of low self-control. *Criminology*, 42(3):773–804, 2004.

D. Nagin. *Group-Based Modeling of Development*. Harvard Univ Pr, 2005.

D. Nagin and J. Waldfogel. The effects of criminality and conviction on the labor market status of young british offenders. *International Review of Law and Economics*, 15(1): 109–126, 1995.

D. S. Nagin. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4:139–157, 1999.

D. S. Nagin and K. C. Land. Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson mode. *Criminology*, 31(3): 327–362, 1993.

D. S. Nagin and C. L. Odgers. Group-based trajectory modeling (nearly) two decades later. *Journal of quantitative criminology*, 26(4):445–453, 2010.

D. S. Nagin and R. E. Tremblay. Developmental trajectory groups: Fact or a useful statistical fiction?*. *Criminology*, 43(4):873–904, 2005.

D. S. Nagin, D. P. Farrington, and T. E. Moffitt. Life-course trajectories of different types of offenders. *Criminology*, 33(1):111–139, 1995.

Ohio State University. Center for Human Resource Research. National longitudinal survey of youth, 1997.

N. Okereafoezeke. Traditional social control in an ethnic society: Law enforcement in a nigerian community 1. *Police Practice and Research*, 4(1):21–33, 2003.

D. P. Phillips and L. L. Carstensen. Clustering of teenage suicides after television news stories about suicide. *New England Journal of Medicine*, 315(11):685–689, 1986.

A. M. Piehl and G. Williams. *Institutional Requirements for Effective Imposition of Fines*. Controlling Crime: Strategies and Tradeoffs. National Bureau of Economic Research, 2011.

A. R. Piquero, D. P. Farrington, and A. Blumstein. The criminal career paradigm. *Crime and justice*, pages 359–506, 2003.

P. Reuter, R. MacCoun, P. Murphy, A. Abrahamse, and B. Simon. *Money from Crime: A Study of the Economics of Drug Dealing in Washington, DC*. RAND, 1990.

R. J. Sampson and J. H. Laub. Crime and deviance over the life course: The salience of adult social bonds. *American Sociological Review*, pages 609–627, 1990.

R. J. Sampson and J. H. Laub. Lifeâcourse desisters? trajectories of crime among delinquent boys followed to age 70*. *Criminology*, 41(3):555–592, 2003.

R. J. Sampson and J. H. Laub. Seductions of method: Rejoinder to nagin and tremblay's developments trajectory groups: Fact or fiction. *Criminology*, 43:905, 2005.

R. J. Sampson, S. W. Raudenbush, and F. Earls. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328):918, 1997.

K. D. Schmidt. *On the Covariance of Monotone Functions of a Random Variable*. Professoren des Inst. fÃ¼r Math. Stochastik, 2003.

T. Skardhamar. Reconsidering the theory on adolescent-limited and life-course persistent anti-social behaviour. *British Journal of Criminology*, 49(6):863–878, 2009.

T. Skardhamar. Distinguishing facts and artifacts in group-based modeling. *Criminology*, 48(1):295–320, 2010.

L. P. Spear. The adolescent brain and age-related behavioral manifestations. *Neuroscience Biobehavioral Reviews*, 24(4):417–463, 2000.

Stata Corporation. *Stata Base Reference Manual: Release 11*. College Station, Texas, 2009.

R. W. Thurston. *Lynching: American Mob Murder in Global Perspective*. Ashgate Pub Co, 2011.

P. E. Tracy, M. E. Wolfgang, and R. M. Figlio. *Delinquency Careers in Two Birth Cohorts*. Plenum Press, 1990.

P. Tremblay and C. Morselli. Patterns in criminal achievement: Wilson and abrahamse revisited. *Criminology*, 38(2):633–657, 2000.

M. E. Walsh. *The Fence: A New Look at the World of Property Theft*. Greenwood Press, 1977.

C. V. Wedgwood. *Thomas Wentworth First Earl of Strafford 1593-1641*. 1962.

D. J. West and D. P. Farrington. *The Deliquent Way of Life*. Heinemann, 1977.

J. Q. Wilson and A. Abrahamse. Does crime pay? *Justice Quarterly*, 9(3):359–377, 1992.

J. Q. Wilson and R. J. Herrnstein. *Crime and Human Nature: The Definitive Study of the Causes of Crime*. Free Press, 1998.

F. E. Zimring. Toward a jurisprudence of youth violence. *Crime and Justice*, 24:477–501, 1998.

Vita

Geoffrey Fain Williams

2007-2012 Ph. D. in Economics, Rutgers University

1997-1999 M. B. A. from The Wharton School, University of Pennsylvania

1988-1991 B. F. A. in Acting from New York University

2010-2012 Research Economist, Princeton Economics Group, Princeton, NJ

2008-2012 Instructor, Rutgers University, Economics and Expository Writing

2004-2007 Associate Director, Africa House, New York University

1994-2004 Manager and Consultant, Internet Services and Cellular Industries