Empirical Research Review and Hypotheses

The theoretical review in the previous chapter sets the stage for a thorough summary of the extant empirical literature concerning the key issues of this study: (1) the relationship between age and crime (and how such a relationship supports or refutes the existence of two discrete groups of offenders and the stability of differences in criminal propensity over time within identifiable groups) and (2) the relationship between past and subsequent criminal activity among individuals in our samples. The two main sections of this chapter review previous studies that have addressed these issues. Included at the end of both sections is a discussion of the general findings, the limitations of the prior research, and the hypotheses that guide this study. This chapter concludes with a discussion of the possible contributions this study can make to the extant literature.

Studies of the Age–Crime Curve

Given the concerns of Blumstein and colleagues (1986, 1988a, 1988b) and Moffitt (1993, 1997) that age–crime curves aggregated over individuals (i.e. calculated for samples as a whole) may conceal considerable heterogeneity in the offending trajectories of individuals, this review is limited to studies in which the authors have disaggregated their samples into ‘latent classes’ or ‘latent groups’ on the basis of the similarity of their longitudinal offending trajectories. Land (1992) has noted that distinguishing between the various age–crime relationship arguments requires the use of models specified at the individual level that specifically allow for incorporating controls for heterogeneity in the propensity to offend. The statistical methods available for modeling the presence of separate trajectories have only become available since Nagin and Land (1993) formulated a statistical model consistent with Land’s (1992) recommendations.
Nagin and Land (1993) introduced the use of semiparametric mixture models to the discipline of criminology as a statistical method able to identify distinct trajectories of criminal offending. Accordingly, all of the studies reviewed below employ the use of the finite mixture methods of Nagin and Land. These finite mixture methods assign each individual to the latent class with the trajectory that most closely resembles the individual’s actual observed crime trajectory. Briefly, the mixture methods of Nagin and Land explicitly assume that the sample (population) is composed of a ‘mixture’ of internally homogenous groups, each with their own distinct trajectory, and this modeling strategy both extracts the underlying trajectories present in the data and assigns each individual to the group to which he/she has the highest posterior probability of belonging (Nagin 1999). Table 3.1 summarizes the key information contained in the individual studies reviewed here. The interested reader is encouraged to examine the original articles for their graphical depictions of the trajectories.

As previously mentioned, Nagin and Land (1993) were the first to present evidence concerning both the number of distinct latent classes or offender groups and to discuss the relationship between age and crime for each specific group. In this influential article, they presented their semiparametric finite mixture Poisson model and applied it to the data from the Cambridge Study in Delinquent Development of West and Farrington (1973, 1977), which is a prospective study of 411 males from a working-class section of London that began in 1961 when the boys were 8 years old.

---

1 In essence, the model fits separate constants and age parameters for each latent class, which allows the shape of each latent class’s trajectory to be distinct.

2 Two studies (Fergusson, Horwood, and Nagin 2000; and Chung et al. 2002) that employ the use of finite mixture models are not reviewed here. Fergusson, Horwood, and Nagin (2000) studied the age–crime curve for a sample of adolescents born in Christchurch, New Zealand, in 1977. Criminal offending data were only available from ages 12 to 18, and the authors note that their study thus presents a very limited view of the age–crime curve because both the childhood and adulthood years were truncated within the analysis. The other study, by Chung et al. (2002), used data from the Seattle Social Development Project (SSDP), a longitudinal study of male and female youths originally drawn from 18 Seattle public elementary schools. The dependent variable in the study consisted of self-report offense seriousness scores (measured at five time points between ages 13 and 21). Since their results do not speak to the issue of trajectories of criminal offending, but rather to trajectories of offense seriousness, the reported results provide ambiguous evidence concerning the relationship between age and crime as used herein.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Data source</th>
<th>Sample size</th>
<th>Gender</th>
<th>Ages studied</th>
<th>Risk level</th>
<th>Dependent variable</th>
<th>Classes found</th>
</tr>
</thead>
<tbody>
<tr>
<td>D’Unger et al. (1998)</td>
<td>1958 Philadelphia Birth Cohort</td>
<td>1,000</td>
<td>Males</td>
<td>8–26</td>
<td>Low</td>
<td>Police contact counts</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1942 Racine Birth Cohort</td>
<td>353</td>
<td>Males</td>
<td>8–30</td>
<td>Low</td>
<td>Police contact counts</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1949 Racine Birth Cohort</td>
<td>721</td>
<td>Males</td>
<td>8–25</td>
<td>Low</td>
<td>Police contact counts</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1955 Racine Birth Cohort</td>
<td>1,067</td>
<td>Males</td>
<td>8–22</td>
<td>Low</td>
<td>Police contact counts</td>
<td>5</td>
</tr>
<tr>
<td>Laub, Nagin, and Sampson (1998)</td>
<td>1950 Glueck Study</td>
<td>480</td>
<td>Males</td>
<td>7–32</td>
<td>High</td>
<td>Arrest counts</td>
<td>4</td>
</tr>
<tr>
<td>Piquero et al. (2001)</td>
<td>California Youth Authority Parolees</td>
<td>272</td>
<td>Males</td>
<td>18–33</td>
<td>High</td>
<td>Arrest counts</td>
<td>6</td>
</tr>
</tbody>
</table>

*Risk level is defined here as 'Low' and 'High'. Low-risk samples correspond to general population samples that are likely to include a majority of low-risk cases in the data. High-risk samples, on the other hand, refers to samples where high-risk cases will constitute the majority of cases (e.g. samples of parolees).
The Nagin and Land study used criminal conviction data gathered between the ages of 10 and 32, with eleven time 'periods' of conviction counts comprising the dependent variables used in the analyses (e.g. convictions at ages 10 and 11 constituted one 'period' of data, 12 and 13 another period, and so on).

Applying the semiparametric mixed Poisson model to these data, Nagin and Land uncovered four distinct groups of offenders. The groups were named according to their offending 'style': 'non-offenders' (64% of the sample), 'adolescent-limited' (12.7%), 'low-rate chronics' (9.9%), and 'high-rate chronics' (13.4%). The non-offenders group was comprised of the sample members who had no convictions during the follow-up period. Obviously, the use of convictions (rather than police contacts or arrests) made it likely that this category would constitute the largest group in the data set.

Importantly, Nagin and Land uncovered three different groups of offenders within these data, each with their own distinct offending trajectory. They noted that there was considerable heterogeneity in the peak age of offending among the various groups. The peak age of offending for the 'adolescent-limited' group was 14, whereas it was 18 and 22, respectively, in the 'high-rate chronics' and 'low-rate chronics' groups. The rate of offending at the peak age (as measured though conviction counts) also varied dramatically among the three groups: with 0.63 convictions for the 'adolescent-limited,' 1.22 convictions for the 'high-rate chronics,' and 0.27 for the 'low-rate chronics.'

Interestingly, their analyses contradicted the assertions of Gottfredson and Hirschi (1990), by finding that between-group age differences in convictions were not stable over time. Although the 'low-rate chronics' group did have a peak offending age, their overall trajectory was amazingly flat between the ages of 16 and 30, and the difference in offending rates between the low- and high-rate chronics groups was only 0.15 by age 30, whereas the difference was about 1.0 at age 16. The 'high-rate chronics' group was already highly active in crime at age 10, with this group already having an average conviction rate of roughly 0.8 convictions at that precocious age. This group did, however, show a significant decrease in their conviction patterns (after their peak rate at age 18) as they progressed through adulthood, a finding that is consistent with the assumptions of both Sampson and Laub (1993) and Gottfredson and Hirschi (1990). By finding a 'low-rate chronics' group, Nagin
and Land were the first researchers to offer empirical evidence of considerably more heterogeneity than the two subgroups posited by Moffitt’s dual taxonomy theory (1993, 1997).

D’Unger and colleagues (D’Unger et al. 1998) conducted the most extensive examination of the age–crime curve to date when they analysed five separate data sets. One set of data pertained to the same set employed in the Nagin and Land (1993) study, and since the results obtained in these two studies are identical, they are not discussed here. The four new sets of results presented by D’Unger and her colleagues include analyses of data from the 1958 Philadelphia Birth Cohort study (Tracy, Wolfgang, and Figlio 1990), and the 1942, 1949, and 1955 Racine Birth Cohort studies (Shannon 1988, 1991).

The 1958 Philadelphia Birth Cohort study longitudinally tracked all 13,160 males and 14,000 females born in Philadelphia in 1958 and who resided in the city from their 10th through their 18th birthday. The frequency of ‘police contact’ for felony and/or misdemeanor criminal offenses was collected through age 26 from Philadelphia Police Department records. Police contacts included both actual arrests by law enforcement personnel as well as law enforcement ‘contacts’ that were handled ‘remedially’ or ‘informally’ (e.g. released at the scene or released to parents) and did not involve a formal arrest where the individual is taken into custody (Tracy and Kempf-Leonard 1996). For computational reasons, D’Unger et al. (1998) estimated their models on a random sample of 1,000 males.3

D’Unger et al. (1998) found a five-class or group model to be the best fit to these data, and named their classes by the nature of the respective offending trajectories of the groups. The largest group was labeled ‘non-offenders’ (comprising 61% of the sample), and the other groups included a ‘high-rate adolescence peaked’ group (1%), a ‘low-rate adolescence peaked’ group (9%), a ‘low-rate chronic’s’ group (21%), and a ‘high-rate chronic’s’ group (8%).4 Interestingly, the four groups who engaged in some level of offending bifurcated into high- and low-rate versions of ‘adolescence peaked’ and ‘chronic’ types that tracked each other fairly well over time.

---

3 Estimation times of finite mixture models tend to increase greatly with sample size (Vermunt and Magidson 2000). Also, to keep the results comparable to those obtained by Nagin and Land (1993), only males were included in these analyses.

4 D’Unger et al. (1998) refer to the group as ‘adolescence peaked’ rather than ‘adolescent-limited’ because their offending patterns included ages outside the adolescent years of 13–17.
Although changes in the rates of offending varied both within and among the ‘adolescence peaked’ and the ‘chronics’ offender groups, each group showed a decrease in offending throughout adulthood. The peak ages of offending were 16 for the ‘adolescence peaked’ groups, and 18 for the chronic groups. The offense rates peaked at 1.0 (for the ‘low-rate adolescence peaked’ group), 3.3 (for the ‘high-rate adolescence peaked’ group), 0.21 (for the ‘low-rate chronics’ group), and 0.95 (for the ‘high-rate chronics’ group). By age 26, however, only the ‘high-rate chronics group’ still had a non-zero offending rate, and their rate at that age was roughly one-quarter of its peak rate at age 18.

The Racine Birth Cohorts longitudinally tracked the offense histories of all individuals born in Racine, Wisconsin, in 1942, 1949, and 1955. For research with these data sets, the dependent variables were the number of police contacts for felony and misdemeanor criminal offenses between the ages 8 and 30 (1942 cohort), ages 8 and 25 (1949 cohort), and ages 8 and 22 (1955 cohort). To make these findings comparable with the previous studies discussed above, the authors limited their analyses to the white and black male members of the samples. This resulted in final sample sizes of 353 (for 1942), 721 (for 1949), and 1,067 (for 1955) individuals, respectively.

For the 1942 birth cohort, D’Unger et al. found the best-fitting model to have five distinct offender groups. These groups included a ‘non-offenders’ group (34.6%), an ‘adolescence peaked’ (20%) group, a ‘low-rate chronics’ (31.4%) group, a ‘high-rate chronics’ group (8.8%), and a ‘late-onset chronics’ group (5.1%). Unlike the findings of previously reviewed studies, one offender group was located in this data set that actually increased their offending with age (the ‘late-onset chronics’ group), with the peak rate of offending for this group occurring at age 28, where it then stabilized through the end of the follow-up period. At age 16, this group had an offending pattern that virtually tracked the ‘adolescence peaked’ group. At that point, however, the two trajectories diverged, with the ‘late-onset chronics’ group continuing to escalate their offending behavior, while the ‘adolescence peaked’ group began to desist from offending. Interestingly, the high- and low-rate chronics groups differed substantially in their offending rates between ages 16 and 22 (with their offending rates differing by about 1.0 police contacts per year). By the end of the follow-up period, however, the offending rates of these two groups were nearly identical.
The ‘high-rate chronic’ group experienced a significant decline in offending in early adulthood, whereas the ‘low-rate chronic’ group was observed to have exhibited a much slower rate of decrease in their offending rate.

A four-class (or group) model for the 1949 sample provided the best fit to this data set. The group trajectories found by D’Unger et al. for this sample included a ‘non-offenders’ group (35%), ‘high-rate chronic’ group (peak age = 18; peak rate = 2.1; 5% of sample), a ‘low-rate adolescence peaked’ group (peak age = 18, peak rate = 0.25; 40% of sample), and a ‘high-rate adolescence peaked’ group (peak age = 18; peak rate = 0.75; 19% of sample). By age 25, both the low-rate and high-rate adolescence peaked groups had virtually desisted from offending (as measured by police contacts), whereas the ‘high-rate chronic’ group was still experiencing roughly 1.5 police contacts per year at this period in their lives. It is interesting to note that this is the only data set for which the trajectories generally followed the proportional changes across time argument proposed by Gottfredson and Hirschi (1990).

For the 1955 cohort, a five-group model was found to provide the most accurate fit to the data. The five groups included a ‘non-offenders’ group (44.5%), an ‘early-onset adolescence peaked’ group (2.2%), a ‘late-onset adolescence peaked’ group (15.4%), ‘low-rate chronic’ group (30.1%), and ‘high-rate chronic’ group (7.8%). Unlike the results from the Philadelphia Birth Cohort data, however, the ‘adolescence peaked’ trajectories did not neatly bifurcate into simply high- and low-rate versions; they differed greatly on their age of onset as well as their rates of offending. Also, the crime trajectories of the various groups did not remain proportional; rather the rate of change of the trajectories varied considerably between groups. For example, at age 8 the ‘early-onset adolescence peaked’ group and the ‘high-rate chronic’ were very similar in offending rates. At age 16 their trajectories differed by about 2 arrests per year, and then by age 22 they were nearly identical again. Similarly, the trajectories of the ‘low-rate chronic’ and the ‘late-onset adolescence peaked’ groups were identical until age 15, at which point the ‘adolescence peaked group’ had a surge in offending, while the offending by the ‘low-rate chronic’ held fairly constant thereafter. By age 22, the ‘late-onset adolescence peaked group’ had decreased their offending back to a level near that of the ‘low-rate chronic’ group.
The results of the studies reviewed here so far have shed doubt on the assertion of Gottfredson and Hirschi that there is a relative stability of between-group differences in offending across time as well as the contention by Moffitt that there are only two discrete groups of offenders in the population. Next, we turn our attention to the first analysis of the longitudinal offending patterns among discrete offender groups within a ‘high-risk’ sample.

Laub, Nagin, and Sampson (1998) conducted an analysis of the longitudinal offending patterns of the 480 delinquent boys from the original Glueck study of the criminal careers of delinquent boys in Boston (Glueck and Glueck 1950, 1968). All 480 boys were white and all had appeared in the Boston Juvenile Court in the late 1930s. The Gluecks followed the boys into adulthood until the age of 32. Sampson and Laub (1990, 1993; Laub and Sampson 1988) subsequently reconstructed these data and put them into machine-readable form, and then used these data to develop and test their theory of informal social control. In their 1998 study, Laub, Nagin, and Sampson used the finite mixture methods of Nagin and Land to ascertain if there were distinct groups of offenders even within this select group of chronically delinquent boys.

For this study, the dependent variable was the count number of arrests in each year between the ages of 7 and 32. The authors found that allowing for four distinct groups (or trajectories) provided the best fit to these data. Since all of the members of this sample were official delinquents, there was obviously no ‘non-offenders’ group in this data set. There was, however, a significant amount of heterogeneity even within this select sample of chronic juvenile delinquents. Further, even though all four of the offender trajectories were very similar in their offending rates up through

---

5 Laub and Sampson (2003; Sampson and Laub 2003) have recently updated the Glueck data to age 70. In the final chapter of this book, we connect our results to their most recent work. Here we briefly note that their most recent analyses resulted in six latent classes and that ‘crime declines with age sooner or later for all offender groups’ (Sampson and Laub 2003: 555).

6 In further analyses, Laub, Nagin, and Sampson (1998) also used the resulting latent class indicators as a method of controlling for persistent individual differences in models testing the crime-preventive benefits of a cohesive marriage. The results of their analyses indicated that, even after controlling for unobserved heterogeneity in criminal propensity, a cohesive marriage was a critical factor in the desistence process. Consistent with their age-graded theory, the benefits of a cohesive marriage accrued gradually over time.
age 13, from that point on there was significant variability in the shape of each group’s crime trajectory.

‘Group 1’ consisted of a high-rate chronic group that had an observed peak offending age of 18 (at about 3 arrests per year), and thereafter their trajectory was relatively constant until they reached their late twenties, when their offending rates began to decline. Only 11 individuals in the sample were assigned to this group. ‘Group 2’ was a more moderate chronic offender group, with a peak offending rate of about 1.2 arrests per year at around age 18. This group comprised about 19 per cent of the sample. The offending rate of ‘Group 2’ was relatively constant during their twenties, and began to decline by the end of the follow-up period (ages 30–32). ‘Group 3’ exhibited an offending pattern very similar to ‘Group 2’ through age 16, but then experienced a significant decline in their offending rate over the remaining age distribution curve. By age 32, this group had a negligible offending rate, whereas ‘Group 2’ was still offending at about 0.8 arrests per year at this age. ‘Group 4’ was the group with the lowest offending rate (peak offending rate was 0.7 at age 16). This group, which comprised about 31 per cent of the sample, also had a rather negligible offending rate (of about 0.1) by age 26, where the rate continued to hover for the remaining six years of the follow-up period. The results of this study should be viewed with caution, however, because the subjects in the Glueck data set were not randomly selected from the population, nor randomly selected from juvenile court cases (see Cohen and Vila 1996). Because the results of this study are based on a matching sample that was drawn by convenience from Boston juvenile court records, the generalizability of the results beyond their obtained sample are uncertain.

Using a national probability sample to avoid possible sampling bias, McDermott and Nagin (2001) studied the self-reported offending patterns of the 835 male respondents in the National Youth Survey. The segment of the age distribution studied ranged from 11 through 24, but fewer than half of the respondents were available for sampling at ages 11–13 and 20–24. Therefore, the lack of available data for estimating these segments of the

---

7 Their analyses covered a period of eight ‘age years’, with the actual ages studied varying between the respondents depending on the age of the respondent at the first wave of the interviews. There were actually only six measurement periods used in the analyses due to unequal spacing of the last interview.
age–crime curve demands caution when interpreting the reported results. The dependent variable for this study was a count of self-reported involvement in rape, auto theft, theft of goods worth more than $50, purchasing stolen property worth more than $50, and breaking and entering.

McDermott and Nagin (2001) found that a three-class model provided the best fit to these data. ‘Group 1’ was engaged in offending from ages 11 through 20 at a relatively constant rate (between 1.0 and 1.3 offenses), at which point their offending patterns were found to decline. The offending rate of ‘Group 1’ peaked at age 15. It should be noted that 748 individuals (or 89% of the sample) in this analysis were assigned to ‘Group 1’, and that most of these individuals reported no criminal activity at each age measured. ‘Group 2,’ comprising about 6 per cent of the sample, had a peak of offending at age 11, with 20 offenses per year. Thereafter, this group showed a significant decline in the offending rate through age 19, at which point the rate leveled off at around 5 offenses per year, which continued through age 24. The offending pattern by ‘Group 3’ was nearly antithetical to the pattern observed for ‘Group 2.’ ‘Group 3,’ which contained 5 per cent of the sample, showed a precipitous increase in their offending from ages 11 through 21, where the offense rate peaked at a rate of 30 offenses per year. The offense rate then declined to 23 offenses per year by age 24.

In still another study, Piquero et al. (2001) present an analysis of the age–crime curve for a sample of high-risk cases. This study involved an analysis of the adult offending patterns of a sample of 272 parolees released from the California Youth Authority (CYA) between 1960 and 1970. This is the same youthful offender correctional system from which we analyzed data in this study, although the data gathered by Piquero et al. pre-date the large increase in violent offending that occurred in the state of California in the 1980s and it did not constitute a random sample of CYA wards. The 272 parolees in the Piquero et al. study were ‘older, had more serious commitment offenses, and/or were more uncooperative in other CYA institutions’ (2001: 57) compared to most of the youths housed in this correctional system. These youthful offenders were paroled at age 18 from the CYA and were then followed for sixteen years until age 33. The dependent variable was a count of arrest events between the ages of 18 and 33. Thus, while their study
concerns a limited segment of the age-crime curve (i.e. adulthood only), it is important for inclusion in our review because the authors found considerable heterogeneity in the adult offending trajectories of this select sample of offenders.

In fact, Piquero et al. (2001) report that a model allowing for six distinct trajectories provided the best fit to these data. Importantly, this study found a six-class model to fit the data with and without controls for ‘exposure time’ (i.e. time not incarcerated), although the authors noted that the scale of the arrests trajectories, especially during their early twenties, was affected by controlling for exposure time. In the nomenclature of criminological research, exposure time is referred to as ‘time on the street’ and is used often to control for the amount of time spent incarcerated during the sample period. When a person is incarcerated, they are denied the opportunity to victimize the general non-institutionalized population simply as a consequence of their isolation or ‘incapacitation’ and not because of any change in their motivation to commit criminal behavior in general.\(^8\)

For the models without a control for exposure time, ‘Latent Class 1’ (18% of the sample) was a group that increased their offending through age 21, which was their peak age of offending (at two offenses per year). This group then decreased their rate of offending through age 33; their offending rate was negligible from age 29 onward. ‘Latent Class 2’ (21% of the sample) displayed a trajectory very similar to ‘Latent Class 1,’ only their offending rate was roughly twice the rate of ‘Latent Class 1.’ Their rate peaked at age 21 and then decreased thereafter, but they still had a positive offending rate at age 33 (at about 0.75 arrests per year). ‘Latent Class 3’ (7% of the sample) peaked their offending during the first year after release (age 18), had a small decrease in their offending rate during the early twenties, and then had a relatively constant rate (at about 3 offenses per year) through the remaining ages in the distribution. ‘Latent Class 4’ displayed a trajectory very similar to ‘Latent Class 3,’ albeit at a lower rate of offending than ‘Latent Class 3,’ from ages 18 though 28. ‘Class 4s’ (18% of the sample) arrest rate also peaked at age 18, then held at a relatively constant

\(^8\) The logic of ‘street time’ is that someone who spends twelve months ‘on the street’ and is arrested one time is very different from an individual who spends one month on the street and is arrested one time.
rate through age 28 at about one offense per year, and then displayed a decreasing offense rate through age 33. The fifth latent class (24% of the sample) also had a peak rate of arrest at age 18 (at two offenses per year), then displayed a decreasing arrest rate through age 25, at which point this rate became stable at about 0.5 arrests per year. Interestingly, the sixth latent class (10%) had a very small offense rate at age 18 (about 0.25 arrests per year), but then essentially desisted entirely over the remaining ages. In other words, this group was able to essentially remain arrest-free after parole. It is interesting to ponder whether this group was comprised of an ‘adolescent-limited’ group who displayed a high offending rate during their adolescent years but then was able to remain arrest-free during adulthood.

Two important differences arose after allowing for an ‘offset’ or control for crime exposure time in the Piquero et al. study. First, the predicted arrest count was of greater magnitude for both ‘Latent Class 2’ and ‘Latent Class 3.’ ‘Latent Class 2’ peaked two years earlier than did ‘Class 3’ at a rate of 7 arrests per year, but then exhibited a pronounced declining arrest rate through age 25, at which point the trajectory assumed the same shape it did in models without controls for ‘time on the street.’ ‘Latent Class 3’ also did not experience its decline in the early twenties, but rather this group’s arrest trajectory held rather constant over most of the remaining age distribution (at a rate of about 7 arrests). While in their early thirties, the group had a decrease in offending of about one arrest per year, for the most part, the overall shape of those two curves did not vary dramatically between the two models. A second difference among the groups noted by Piquero et al. (2001) was that ‘Latent Class 4’ also did not experience a decline in offending in the early twenties or a further decline in the thirties, but rather this group had a constant arrest rate over the entire age distribution (at about three arrests per year). The remaining three classes had trajectories that were essentially identical in both models. Piquero et al. (2001: 68) concluded that ‘the general shape of the arrest trend appears to be robust to controls for exposure time.’ The percentage of cases assigned to each latent class was virtually identical across the two models as well. Piquero and his colleagues report that more than anything else, it was the magnitude of the arrests scale that was affected the most by controlling for street or exposure time. Nonetheless, the results from both of
their models indicate that there is significant heterogeneity in the adult arrest patterns of these serious offenders, but how that adulthood heterogeneity related to prior existing differences could not be determined with these data. Indeed, the results of this study also leave one wondering if the findings would have changed had they access to either the juvenile arrest histories of their sample members or to a much larger, random sample of youthful offenders. Our point is that with an overall sample size of only 272 cases that were not randomly drawn, the generalizability of the findings from this study must be viewed cautiously.

We have now completed our review of the prior studies that have addressed the age–crime relationship within discrete offender groups (that are internally homogeneous with respect to their offending patterns across time). In the next section, we place the results of these studies into perspective with a discussion that focuses on both the significant themes that have emerged and the methodological limitations of these prior studies. We then specify four of the eight hypotheses to be tested in our study.

**Discussion and Hypotheses Related to the Age–Crime Relationship**

A general summary of many of the studies we have reviewed here can be found in the first, and arguably definitive, study concerning the relationship between age and crime within distinct offender groups by Nagin and Land. In this study, Nagin and Land (1993: 358) noted, `our findings point to large variation across the population not only in offending levels by age but also in the trajectory of offending over age.' The results of their study are illustrative of several themes in the literature particularly relevant to the study we will conduct herein.

First, there appears to be a considerable amount of individual variation in the offending rates of individuals. This heterogeneity in offending propensity has been documented across a variety of different settings, including birth cohorts from a small Midwestern town such as Racine, to a large urban city such as Philadelphia (D’Unger et al. 1998), in a random sample of the general population (McDermott and Nagin 2001) to samples of the serious youthful offender population from Boston and California (Laub, Nagin, and Sampson 1998; Piquero et al. 2001), in a nationwide sample that
uses self-report data (McDermott and Nagin 2001) to samples that use official data (Nagin and Land 1993; Piquero et al. 2001), and across varying cultural settings such as England (Nagin and Land 1993) and New Zealand (Fergusson, Horwood, and Nagin 2000). The generalizability of the finding of heterogeneity in individual offending is extremely important because samples are often treated as if one trajectory or group is present in the data and as if the effects of persistent heterogeneity are trivial. Such short-sightedness by researchers can lead to misleading and erroneous conclusions (D’Unger et al. 1998; Land, Nagin, and McCall 2001; Maltz 1994; Moffitt 1993, 1997).

A second theme in this literature concerns the significant amount of between-group heterogeneity displayed with respect to the direction and nature of change in the shape of individual’s crime trajectories across the age distribution. Given that the theories of Gottfredson and Hirschi (1990) and Moffitt (1993, 1997) predict near, clearly defined changes in offending trajectories over time, while Sampson and Laub (1993) predict more heterogeneity in crime trajectories over time (especially in the adult years), the evidence at this point would appear to lend more support for the theoretical position of Sampson and Laub. For example, the McDermott and Nagin (2001) study found a crime trajectory that continued to increase across age, while D’Unger et al. (1998) discovered high- and low-rate ‘chronics’ display markedly slower change in their crime trajectories in comparison to the adolescence peaked groups in their data. Laub, Nagin, and Sampson (1998) report crime trajectories in their sample that were quite similar at early ages to show markedly differential growth patterns during adulthood.

Still a third theme contained in this literature involves two trajectory ‘regularities’ in many of the studies reviewed here. The first regularity is that two distinct primary groups have been uncovered across many of the studies: (1) the ‘chronic offender’ group where crime peaks between ages 17–21 and then drops slowly in the remaining years that they are in their twenties and (2) the ‘adolescent peaked’ group where crime peaks between ages 15–18 and then drops rapidly to near zero by age 22 (D’Unger et al. 1998). The second common pattern or regularity within these studies is a common crime trajectory shape that often bifurcates into high- and low-rate groups that track each other over the age distribution (D’Unger et al. 1998).
Yet, even in the face of these regularities, it should be noted that neither the longitudinal shapes nor the number of distinct trajectories were entirely consistent across the various studies discussed above. For example, several crime trajectory patterns have only been identified in one or two of these studies, most notably the ‘late-onset chronic’ offender found in D’Unger et al.’s (1998) analysis of the 1942 Racine Birth Cohort. While all of the aforementioned studies uncovered more than two discrete groups, the exact number of classes has ranged from three in the McDermott and Nagin (2001) study to six in the Piquero et al. (2001) study. Most studies report the identification of four or five distinct crime trajectories among their data. Notably this finding directly contradicts Moffitt’s hypothesis of two distinct offender trajectories and suggests that more than two trajectories are needed to capture the variation of offending trajectories in the population. If there are not just two distinct offender trajectories, then how many are sufficient to accurately reflect existing crime patterns over the age distribution? Does the number of crime trajectories identified depend on the sample composition? How stable are the identified latent classes within a given population over time? While definitive answers to these questions will require much future research, results such as those we present in Chapter 7 can expand our understanding of these issues.

The current literature of the age–crime curve for distinct groups of offenders has several limitations that highlight the need for further study. First, some of the previous studies have focused on rather limited segments of the age distribution (due to limitations of the data sets), with several studies not beginning their measurement of offending behaviors until the onset of adolescence or later, and most of the studies ending their follow-up periods prior to or around the age of 25. The study of Laub, Nagin, and Sampson (1998) has the longest follow-up period to date, examining the nature of the offending trajectories of 480 delinquents from age 7 through age 32. The nature and shape of offending patterns beyond the early thirties are currently not well understood. Second, the study of D’Unger et al. (1998) is the only study to compare the results from data sets generalizable to the same population over time. This makes it very difficult to replicate not only the existence of a crime trajectory group over time, but also whether there are any changes in the precise number or nature of the offending trajectories over time. As such, D’Unger et al. (1998) argue that
replication of offending trajectories is a critical research need that is necessary to prevent reifying any particular identified offending trajectory as a stable element in a population. As D’Unger et al. note (1998: 1624–5):

The effects of age, cohort or sample composition, and historical setting all play important roles in influencing individual development, hence the variation in trajectories over time. Social context must be viewed as a ‘force in development’ (Elder and O’Rand 1995), which has the power to alter trajectories of myriad types of behaviour.

A final limitation of the previous research reviewed here concerns the analyses of the ‘high-risk’ samples; only two studies have focused on select samples of ‘high-risk’ offenders. Both of those studies, however, have limitations that require additional research on this critical segment of the offending population. The Laub, Nagin, and Sampson (1998) study was based on the offending patterns of a non-random sample of white, male delinquents from Boston measured from the 1930s to the 1960s, and thus a key question is whether trajectories similar to the ones they describe can be found in more contemporary samples of the population. This is especially significant given that the nature of criminal offending appears to have changed dramatically (i.e. became more violent) after the point in time when Laub, Nagin, and Sampson’s data were gathered. Piquero et al. (2001), on the other hand, only had access to data regarding offending patterns of a sample of serious youthful offenders during their adult years (age 18–33). Data from these subject’s juvenile years were entirely absent from the analyses. This limitation raises several interesting questions with respect to this age segment of the population: (1) how do differences in offending trajectories during the juvenile years relate to the nature of offending during the adult years; and (2) is there an adolescent-peaked group within this population?9 Furthermore, both of these studies were based on comparably ‘small samples,’ and thus we wonder to what degree their identification (of a particular latent class) is a consequence of sampling variation? Again, this question becomes more interesting once we consider that neither of the samples used in these two studies was randomly drawn. Thus, it is our contention that there is a critical need for subsequent empirical

9 Recall that Piquero et al. (2001) found a group with an offending trajectory that by age 20 had terminated their criminal activity (as measured by arrests).
investigations of the nature of offending trajectories with random samples from the population of serious youthful offenders, a contention that has been echoed by Laub and Sampson (2001), Scholte (1999), and Tolan and Gorman-Smith (1998).

Given the findings and limitations of the literature discussed above, our study will investigate four hypotheses related to the age–crime curve using three relatively large, random samples of serious youthful offenders (to be described in greater detail in Chapter 5):

**H1:** There are multiple groups or latent classes of offenders with distinct arrest trajectories even on the high end of the criminal propensity continuum where the serious youthful offenders are located.

**H2:** There are more than two groups of offenders with distinctly different arrest trajectories even on the high end of the criminal propensity continuum.

**H3:** There is an adolescence-peaked group even in samples of serious youthful offenders.

**H4:** The age–crime curve is invariant among the latent classes of serious youthful offenders. Between-group differences will not vary across time.

These hypotheses are largely based on both the prior empirical results from the Laub, Nagin, and Sampson (1998) and Piquero et al. (2001) studies that indicate there is a significant level of heterogeneity in the offending patterns of serious youthful offenders, as well as the theoretical arguments of Cohen and Vila (1996) and D’Unger et al. (1998) that hypothesize the possibility of greater heterogeneity on the far right tail of the crime continuum than previously suggested. **H2** and **H3** are central to Moffitt’s (1993, 1997) theoretical perspective, while **H4** is central to the theories of Gottfredson and Hirschi (1990) and Sampson and Laub (1993).

Evidence supporting **H2** would cast doubt on the adequacy of Moffitt’s theory that there are only two offender groups in the population. Evidence supporting **H4** would support Gottfredson and Hirschi (1990), while evidence refuting it would support the theoretical position of Sampson and Laub (1993).

Findings in support of these hypotheses are important for the crime literature because serious youthful offenders are often referred to as being ‘relatively homogeneous’ (Ge, Donnellan, and Wenk 2001: 750). As a whole, serious youthful offenders are an elusive class of offenders because they are (fortunately) relatively rare in the population of offenders (Cernkovich, Giordano, and Pugh 1985). Researchers thus are often forced to empirically ‘lump’ together
offenders who have met some minimum definitional criteria that usually involve a measure of either seriousness and/or chronicity of offending (McDermott 1983). After ‘making the cut,’ this group of offenders is usually isolated and treated as a homogeneous group (often labeled as the ‘chronic offender’ group).\(^\text{10}\) If there is significant heterogeneity in the propensity to offend within this group of offenders, recognition of that fact is important to the crime literature for both its theoretical and public policy implications.\(^\text{11}\)

The results supporting or disconfirming these four hypotheses will be presented in Chapter 7. In that chapter, we will apply Nagin and Land’s (1993) semiparametric mixed Poisson model to each of the three samples employed in our study. After determining the optimal number of latent classes of serious offenders present in each sample, the arrest trajectories will be graphed over the age distribution. Comparisons of the trajectories will be made concerning the patterns of offending displayed over time within and between the latent classes.

**Studies of the Relationship of Past to Subsequent Criminal Activity**

We now turn our attention to reviewing previous studies of the second critical issue addressed in this study—the relationship between past and subsequent criminal activity. Since investigating

\(^\text{10}\) Loeber, Farrington, and Waschbusch (1998) provide an extensive discussion of the variable cut-off points that have been used in an attempt to isolate the type of offenders described in our study.

\(^\text{11}\) For example, Ge, Donnellan, and Wenk (2001) analyse the arrest patterns of a sample of 2,263 males committed to the CYA in 1964 and 1965. The authors analyse the arrest patterns of the CYA wards at ages 18–20, 21–25, 26–30, and over 31 using a series of ordinary least squares regression models. The authors conclude (from a state dependence position) that ‘early problem behaviors exert a significant influence on persistent offending. Early involvement with alcohol and drug use was a significant predictor of adult arrest frequency. This suggests that early substance use and abuse can influence criminal behavior throughout the life span.’ Their analyses did not control for unobserved heterogeneity, however, and thus it could simply be that early drug use and abuse is correlated with the unmeasured (or at least the poorly measured) heterogeneity in the propensity to offend. Lacking controls for unmeasured heterogeneity, this finding is subject to criticism as a methodological artifact that would disappear if one were to use a more appropriate statistical model. See the section below describing studies of the relationship between past and subsequent criminal activity for a further discussion of the importance of unmeasured heterogeneity.
the relationship between past and subsequent criminal activity requires longitudinal (panel) data on a set of individuals, the studies reviewed here are limited to those studies following a panel of individuals over time. Furthermore, given the differential explanations of the population heterogeneity and state dependence positions for the underlying causes of the correlation between criminal activity at two different points in time, all of the studies reviewed here also control for individual differences in the propensity to commit criminal acts.

Historically, controlling for differences in criminal propensity has been most often attempted by including control variables measuring individual characteristics or other factors considered relevant in a regression model. However, multiple studies (Bushway, Brame, and Paternoster 1999; Nagin and Paternoster 2000; Paternoster and Brame 1997; Paternoster, Brame, and Farrington 2001) argue that there are two principal problems with this method of controlling for individual differences. First, criminologists cannot agree on the precise and most appropriate measures or variables that reasonably capture individual differences in criminal propensity. Second, even if there was such a consensus on relevant variables, most data sets probably would not have some, most, or perhaps any of those key measures.

The end result of such problems, as noted by Nagin and Paternoster (2000: 131), is that ‘researchers would have no way of knowing if they have captured a sizeable share of the between-individual variation in criminal propensity with the measures they have available to them. Consequently, perhaps the lion’s share of criminal propensity would be unmeasured or unobserved.’ Simulations by Bushway, Brame, and Paternoster (1999) show that failing to account for unobserved heterogeneity leads to seriously biased estimates that favor the state dependence argument (see also Heckman 1981c; Hsiao 1986). Unobserved heterogeneity is, in essence, akin to omitting a key variable from the model specification, resulting in biased estimates of the other included covariates that are correlated with the omitted variable (Bushway, Brame, and Paternoster 1999; Nagin and Paternoster 2000). Since prior criminal activity will be positively correlated with criminal propensity, failure to adequately control for persistent unobserved heterogeneity (in criminal propensity among individuals) will lead to positively biased coefficients for the variable representing prior
criminal activity. In other words, without statistically controlling for persistent individual differences, the coefficient for the prior criminal activity variable will absorb the effect of the omitted variable (individual differences in criminal propensity), resulting in an overestimate of the effect of prior criminal activity on present criminal activity. As Nagin and Paternoster (2000: 131) explicate, ‘if there is unobserved heterogeneity that accounts for offending over time, the failure to explicitly consider this will lead to biased estimates of observed time-varying factors.’

Thus, all of the studies to be reviewed below make use of statistical techniques controlling for unobserved or ‘hidden’ heterogeneity. These studies have used one of two statistical methods (and in one case both methods) to account for unobserved heterogeneity: (1) parametric random effects statistical models or (2) semiparametric random effects statistical models. The primary difference between the two methods concerns the distribution of the unobserved heterogeneity. The parametric random effects models assume that the unobserved heterogeneity is continuously distributed in the population according to some known (mathematically tractable) parametric distributional form (e.g. it is normally distributed). The semiparametric form of the models nonparametrically approximate the form of the unobserved heterogeneity, assuming only that an approximation can be accomplished using a discrete, multinomial distribution (Heckman and Singer 1984; Land, McCall, and Nagin 1996; Nagin and Land 1993). This semiparametric random effects model is, in fact, the finite mixture model of Nagin and Land (1993) previously discussed. Each ‘latent

---

12 For readers desiring more information on these models at this point in their reading, the models are described in Chapter 5. Briefly, the parametric random effects model assumes that the error term for an individual at any point in time is composed of two components: a time-invariant, individual-specific term and a pure random disturbance term (that is distributed according to some parametric assumption, usually the normal distribution). The individual-specific component, which is invariant over time, is presumed to capture persistent, unmeasured individual differences in the propensity to offend. The correlation of an individual’s error term over time (referred to as rho or ρ) is calculated as the variance of the individual-specific terms divided by the variance of the total error term (Hsiao 1986; Nagin and Paternoster 1991). Rho (ρ) is an estimate of the proportion of the variance of the error term that is due to persistent (time-stable) heterogeneity. If ρ = 1, the variance of the error term is entirely due to heterogeneity, whereas if ρ = 0, then persistent heterogeneity is negligible (Nagin and Paternoster 1991).
class’ is assumed to be a single ‘point-of-support’ or ‘segment’ of the multinomial distribution, and the distribution of unobserved heterogeneity (known as the mixing distribution in statistics jargon) is approximated using a finite number of points-of-support. Within each ‘segment’ of the sample, individuals are internally homogeneous with respect to criminal propensity, but individuals from different segments have varying propensities to engage in criminal activities.

As Bushway, Brame, and Paternoster (1999) note, both of these models make assumptions, and the degree to which the assumptions are tenable is key to the robustness of any observed results. The parametric form of the models is more restrictive and more efficient than is the semiparametric form, which is less restrictive and hence also less efficient. Violations of the assumptions of each model, including the assumption of the distribution of unobserved heterogeneity, can have a significant impact on the conclusions based upon the results obtained from each model. We will return to this significant issue of ‘violating assumptions’ later in our discussion of the studies that address the relationship between past and subsequent criminal activity.

First, however, the results of several studies will be reviewed as they were reported in the original articles. In the following discussion, we try to stay substantively focused, but will include methodologically technical comments and footnotes when necessary. It should be stated that modeling the relationship between past and subsequent criminal activity is methodologically complicated, a point that should not be underemphasized (Nagin and Paternoster 2000). The old adage, ‘the devil is in the details,’ is quite appropriate for this issue. Table 3.2 presents key information from the different studies to be reviewed below that examine the relationship between past and subsequent criminal activity.

One of the key studies regarding the relationship between past and subsequent criminal activity was that of Nagin and Paternoster (1991). It is a study of 1,163 respondents in a convenience sample of students from nine high schools in South Carolina. This panel began in 1981 and had subsequent ‘follow-up waves’ in 1982 and 1983. Although there were initially 2,700 sophomore respondents, at the final wave only 1,250 senior respondents remained in the panel, and of these, only 1,163 filled out the information on relevant covariates deemed necessary for inclusion in the study by the authors. The dependent variable was constructed as a binary variable representing self-reported participation in three types of
<table>
<thead>
<tr>
<th>Authors</th>
<th>Data source</th>
<th>Sample size</th>
<th>Gender</th>
<th>Periods studied</th>
<th>Risk level$^a$</th>
<th>Dependent variable</th>
<th>Finding$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nagin and Paternoster (1991)</td>
<td>1981 South Carolina High School Survey</td>
<td>1,163</td>
<td>Males/Females</td>
<td>2</td>
<td>Low</td>
<td>Binary indicator of self-reported property crimes</td>
<td>SD</td>
</tr>
<tr>
<td>Paternoster and Brame (1997)</td>
<td>1977 National Youth Survey</td>
<td>479</td>
<td>Males/Females</td>
<td>5</td>
<td>Low</td>
<td>Binary and count indicators of self-reported delinquency</td>
<td>Mixed</td>
</tr>
</tbody>
</table>

$^a$Risk level is defined here as ‘Low’ and ‘High’. Low-risk samples correspond to general population samples that are likely to include a majority of low-risk cases in the data. High-risk samples, on the other hand, refer to samples where high-risk cases will constitute the majority of cases (e.g. samples of parolees).

$^b$Finding: SD = state dependence; PH = population heterogeneity; Mixed = both state dependence and population heterogeneity.
property crimes: stealing something valued less than $10, stealing something valued between $10 and $50, and breaking into a building and stealing something. Most respondents who indicated they had participated in one of these crimes had only stolen something valued at less than $10. The panel assessed the dependent variable at two different points in time: participation between waves 1 and 2, and participation between waves 2 and 3. As the authors point out, the length of this panel (2 points in time) is the absolute minimum number of periods needed to estimate a panel model. In their model, the lagged dependent variable (participation in crime during the prior measured period) is the parameter estimate providing evidence for or against a process of state dependence.

Using a (parametric) random effects probit model, Nagin and Paternoster found a significant correlation between participation in the property crimes at the two points in time, even after accounting for unobserved heterogeneity through the use of the random effects model. Participating in the property crime between waves 1 and 2 significantly increased the odds of participating in crime between waves 2 and 3, net of persistent heterogeneity (rho was estimated to be equal to 0, indicating that persistent heterogeneity was negligible). According to Nagin and Paternoster (1991: 183), ‘the results revealed that prior participation in crime had a positive and significant association with subsequent participation, controlling for the possibility of unobserved heterogeneity.’ This finding is consistent with the hypothesis that prior participation reduces the barrier against subsequent participation in crime. The authors were notably cautious in their conclusions, noting that it was an ‘exploratory’ study with ‘suggestive results’, because of several methodological limitations including the use of a non-random convenience sample, the built-in assumptions of the random effects model concerning the distribution of the heterogeneity (i.e. that it was normally distributed), and because the ‘initial conditions’ assumption of the model was clearly violated.\textsuperscript{13} Furthermore, due

\textsuperscript{13} The initial conditions assumption refers to the assumption that, at the first wave of the study, none of the respondents had already initiated the process (i.e. been involved in property crime activity). This assumption is required so that the model is able to obtain an unbiased estimate of the individual-specific component of the error term. It turns out that this assumption is fundamentally critical to calculating unbiased estimates concerning the relationship between past and subsequent criminal activity (Brame, Bushway, and Paternoster 1999).
to heavy sample attrition (57% dropped out of the panel before the final wave), the potential ‘homogenization’ of the sample with respect to criminal propensity could not be ruled out.

In addition to presenting some initial findings on the relationship of past to subsequent criminal activity, the study by Nagin and Paternoster (1991) was also noteworthy because (1) it was the first study to explicitly address and elaborate the state dependence versus population heterogeneity arguments for crime, (2) they proposed the use of the random effects models as a viable method for addressing the issues surrounding the continuity in offending patterns, and, perhaps more importantly, because (3) their findings were so provocative as to stimulate continued research on these issues.

Soon after the publication of the Nagin and Paternoster (1991) study, Nagin and Farrington (1992a) presented results bearing on the issue of the relationship between prior and subsequent crime by using the data from the Cambridge Study in Delinquent Development previously described. Recall that this study employs twenty-two years’ worth of conviction data covering the ages of 10–32 for 403 London-born males. The dependent variable for each period was a binary indicator of conviction during consecutive two-year periods (e.g. any conviction during ages 10 and 11 constituted the offense or dependent variable for the first period). Following the lead of Nagin and Paternoster (1991), Nagin and Farrington (1992a) also used the random effects probit model that assumes unobserved heterogeneity to be normally distributed. Interest focused on the parameter estimate for the binary variable that indicated whether or not the individual had been convicted in the prior period (i.e. this is the lagged dependent variable). The coefficient for that variable represents the estimate of the state dependence effect for these data.

In contrast to the findings of Nagin and Paternoster (1991), this study found a highly significant, strong effect of persistent unobserved heterogeneity that served to significantly reduce the association between past and subsequent criminal activity. In the model ignoring persistent heterogeneity (i.e. a standard probit model), the parameter estimate relating conviction in the prior period to conviction in the subsequent measured period was 1.16. In the model controlling for persistent heterogeneity, the estimate was reduced in magnitude to 0.446, roughly a 62 per cent reduction in
the magnitude of the effect. Rho (the within-individual correlation of the error term across time) was estimated to be 0.4, indicating that 40 per cent of the unexplained error variance was estimated to be due to persistent hidden heterogeneity. Nagin and Farrington (1992a: 253) focused their attention on the reduction of the magnitude of the state dependence parameter after controlling for unobserved heterogeneity and the large magnitude of the rho estimate. They concluded that the results were most consistent with the population heterogeneity position and that ‘evidence of true state dependence is limited. After controlling for persistent unobserved heterogeneity, the association between past and subsequent participation is greatly diminished.’ However, closer examination of their results more clearly supports a ‘mixed’ model where both population heterogeneity and state dependence processes are at work (Paternoster et al. 1997). Nonetheless, this study is important because it demonstrates that population heterogeneity, if left uncontrolled, could have serious effects on the estimates of variables indicating evidence in support of the state dependence process.

Although not explicitly addressing the relationship between past and subsequent criminal activity, further analyses of the Cambridge data by Nagin and Farrington (1992b) also revealed strong effects of persistent unobserved heterogeneity. Employing the same data and statistical models from the previous study, Nagin and Farrington (1992b) investigated whether age of onset had a significant effect on the probability of conviction in the eleven-period panel, net of the effects of persistent unobserved heterogeneity. The state dependence interpretation of the age of onset variable is that early conviction has a significant causal impact on the probability of subsequent criminal activity (i.e. conviction causes changes in their life circumstances, such as increasing the likelihood of association with delinquent peers or reducing the social bond, that makes continuing in a life of crime more likely), whereas the population heterogeneity interpretation is that the age of onset variable is merely a proxy measure indicating the level of criminal propensity (i.e. individuals with an earlier age of onset have the highest criminal propensity levels). In the model without controls for persistent unobserved heterogeneity, the age of onset variable was found to have a large and highly significant negative effect, indicating that as the age at first conviction increased, the
odds of a subsequent conviction decreased. However, in the model controlling for unobserved heterogeneity (i.e. in the random effects probit model), the inverse association between age at first conviction and the odds of a subsequent conviction was reduced to insignificance and near zero in absolute magnitude. In other words, the inverse association between prior and subsequent criminal activity was entirely attributable to the effects of persistent unobserved heterogeneity (i.e. due to time-stable differences).

In an empirical test of their age-graded social control theory, Sampson and Laub (1993) presented a two-period panel analysis of arrest counts between ages 25–32 (period 1) and ages 32–45 (period 2). These analyses employed the use of the Glueck data described earlier, plus results from the matched ‘control group.’ Here ‘non-delinquent’ cases were matched (case by case) to their ‘delinquent’ pairs on the basis of age, ethnicity (e.g. Irish, Italian, German, Jewish), and neighbourhood (n = 289 for the delinquents; n = 401 for the matched control group). Sampson and Laub utilized a generalized least squares (GLS) random effects model (which is in essence a random effects OLS model) and they found results consistent with those of Nagin and Farrington (1992a).

In the models estimated on both the delinquents and the control group, Sampson and Laub found significant levels of persistent unobserved heterogeneity (rho = 0.22 and 0.29 for the delinquents and control groups, respectively). Furthermore, they found that even after controlling for persistent unobserved heterogeneity, both the unofficial and official juvenile delinquency behavioral variables were positively and significantly related to observed crime frequencies between ages 25 and 45. These results were found in the models for both the delinquent group and the control group. Similarly, the results of their analyses also indicated that several other variables representing the state dependence position were

---

14 The actual specification employed the use of two variables to capture the effect of age of onset. One variable indicated if the individual had ever been convicted in a prior period, while the other variable indicated the actual age of onset. The use of two variables allowed the state dependence effect to decrease as the age of onset increased (i.e. it allowed the positive impact of the first variable to be magnified by an early age of onset).

15 Sampson and Laub (1993) present sample sizes for the ‘pooled’ data sets. Since their analyses are based on a two-period panel model, we have divided the pooled sample sizes by two to arrive at the sample size.
significantly related to engaging in crime during adulthood, even after controlling for persistent individual differences. For example, the models for both the ‘experimental and control’ groups indicated that job stability had significant negative effects on adult crime frequency. This suggested that individuals with less stable job histories were more likely to engage in crime during adulthood, net of the effects of persistent individual differences and measures of juvenile offending frequency. According to Sampson and Laub (1993: 198–9), ‘these findings support the idea that state dependence underlies the effects of both prior crime and weak social bonds.’ When all the evidence presented by Sampson and Laub (1993) is considered together, however, their data clearly support the ‘mixed’ position—both population heterogeneity and state dependence processes were found to be present within the Glueck data.\(^{16}\)

In a re-analysis of the Cambridge data used in their initial (1993) article, Land and Nagin (1996) present further evidence concerning the link between having a prior conviction (at any point in the individual’s past) and the probability of a subsequent conviction at a given age. Employing their semiparametric finite (Poisson) mixture model, Land and Nagin (1996) find evidence to support the mixed position that dovetails squarely with the conclusions of Nagin and Farrington (1992a), who had analyzed the same data with the parametric random effects probit model.\(^{17}\) The analyses by

\(^{16}\) Sampson and Laub (1996) analyzed the military arrest history of the samples and also found that early entry into the military significantly improved the lives of the structurally disadvantaged and delinquent men. Thus, military service appeared to be a ‘turning point’ in the lives of these men, allowing them to enrich their lives, including their occupational status, job stability, and socio-economic achievement in adulthood. This effect was especially pronounced among the veterans previously stigmatized as official delinquents. In other words, events during adulthood have important consequences for subsequent outcomes. It should be noted, however, that Laub and Sampson (1998) also present evidence in support of the population heterogeneity argument. In these analyses, Laub and Sampson (1998) found that the delinquent group was significantly less likely to take advantage of educational opportunities available both while in the military and through the G.I. educational bill, and that the chronic offenders were significantly less likely to achieve good marriages.

\(^{17}\) Technically speaking, Land and Nagin (1996) estimated a multiple-spell discrete-time hazard model of the time until conviction (i.e. years until or years between convictions). Land and Nagin (1996) show that under regular conditions, the micro-level Poisson model is equivalent to a discrete-time hazard model. See Land, Nagin, and McCall (2001) for further information regarding this event history formulation of the finite mixture model.
Land and Nagin (1996) were the first to use the finite mixture models (allowing for a nonparametric specification of the distribution of unobserved heterogeneity) to address the question of whether past evidence of engaging in criminal activity has a significant effect on subsequent criminality after controlling for unobserved heterogeneity. Consistent with the conclusions of their initial (1993) article, Land and Nagin (1996) found significant differences in the propensity to offend within the sample. They uncovered the same four distinct trajectory patterns reported in their earlier article. Land and Nagin again uncovered four distinct trajectory patterns as in their earlier article, and the probability of surviving without a conviction varied dramatically across the groups.

Land and Nagin also included a variable in their models to assess whether or not the individual had ever been previously convicted. The parameter estimate for the prior conviction variable indicated that individuals with a prior conviction had a higher ‘hazard’ of being convicted at a given age. Notably, this parameter estimate was calculated net of the effects of unobserved individual differences that were captured through the use of the points-of-support approach to estimate unmeasured heterogeneity in the offending population. Stated differently, if you compared an ‘unconvicted’ individual with an already ‘convicted’ individual within the same ‘segment’ or ‘point-of-support’ of the unmeasured heterogeneity distribution and at the same age, the individual who had been convicted at a prior age had a much greater chance of being convicted at that age. For example, at age 16 the probability of ‘onset’ (or first conviction) within the ‘high-level chronic’ group was 0.321, whereas the ‘post-onset’ probability of conviction within this group was 0.695.

It is worth noting that in their initial article, Nagin and Land (1993) also modeled a state dependence variable (lagged indicator of conviction in the prior period), but they did so within the ‘intermittency’ portion of their model that only included controls for observed heterogeneity (through the inclusion of observed variables) rather than unobserved heterogeneity. The intermittency concept allows for the possibility that periods of criminal activity

---

18 The technical aspect of the intermittency component of their model is that it permits their semiparametric mixed Poisson model to be generalized as a ‘zero inflated Poisson’ model, which allows for more zeroes than would be expected by the standard Poisson distribution (Lambert 1992; Mullahy 1986; Zorn 1998).
may be interspersed with periods of inactivity, yet this inactivity does not signal the end of an individual’s ‘criminal career.’ This portion of the model substantively investigates the factors that predict the probability of being an ‘active’ offender at a given age. Nagin and Land specified this component of the model to be predicted by age, age squared, a lagged indicator of conviction in the prior period, and a composite, time-stable measure called TOT (that was composed of measures of risk-taking attitude, parental criminality, a poor parenting indicator, and IQ). The parameter estimate for the lagged conviction indicator was significantly and positively related to the probability of being an active offender at a given age. It is interesting to note that the parameter estimate (1.09) for the lagged conviction indicator was nearly identical to the parameter estimate that Nagin and Farrington (1992a) found (1.16) in their standard probit model that did not account for unobserved heterogeneity. The intermittency component of the Nagin and Land (1993) analyses was important, since it explicitly demonstrated the presence of within-individual variation in criminal offending, but the authors noted that the theoretical importance of the concept of intermittency was problematic.

Subsequently, however, Horney, Osgood, and Marshall (1995) proposed an explanation for the periods of intermittent offending noted in the Nagin and Land study. These authors connected the possibility of periods of activity as being interspersed with periods of inactivity by drawing on the theory of Sampson and Laub. They argued that the ‘intermittency’ effect could be explained by short-term changes in ‘local life circumstances.’ Adopting a clear ‘mixed’ position, they noted (Horney, Osgood, and Marshall 1995: 658–9):

Although a persistent underlying trait like self-control can influence both an individual’s overall level of offending and his or her stability of marriage and employment, that shared influence does not mean that a relationship between offending and the life circumstances is necessarily spurious. It is still possible that involvement in those social institutions influences the likelihood of offending during the time of involvement. The high crime rate of the most persistent offender, rather than indicating a total lack of investment in social institutions, may instead reflect alternating periods of criminal activity and inactivity.

Using data on short-term variations in both ‘social bonding’ variables (such as going to school, working, living with a wife, drinking
heavily, using drugs) and short-term variations of offending behaviors among a sample of incarcerated prisoners, the results of Horney, Osgood, and Marshall’s (1995) hierarchical linear models analysis showed that short-term, within-individual changes in offending behavior were strongly related to changes in the local life circumstances of the offenders, net of controls for unmeasured heterogeneity in the propensity to offend. The men in this sample (600 serious adult offenders) were significantly less likely to be involved in criminal activity when they were working, were not using drugs or alcohol, and were living with their wives. This finding is entirely consistent with a ‘state dependence’ position and clearly highlights that short-term change in the offenders’ criminal activity is intrinsically related to short-term improvement (or worsening) of their local life circumstances. The implication here is that if criminal arrest/conviction ‘worsens’ the local life circumstances of offenders through its ‘negative effects’ on the odds of obtaining a job, going to school, or living with a wife, then (even after controlling for individual criminal propensity) a strong association between previous and current offending is to be expected.

A more recent study concerning unobserved heterogeneity by Paternoster et al. (1997) followed the example of Sampson and Laub (1993) by examining the offending patterns of a sample of high-risk youthful offenders. Using a sample of 838 young, male offenders released from the training schools of the North Carolina Division of Youth Services in 1988–9, Paternoster et al. (1997) examined the yearly arrest counts of the offenders between the date of release and November 1994, when the follow-up period ended. Using the random effects negative binomial panel model, the results of their analyses are based on four to six years of arrest counts (i.e. an unbalanced panel) in the post-release period. Here the unobserved heterogeneity was assumed to be distributed according to the beta distribution.

Similar to Nagin and Farrington’s study (1992a), Paternoster and his colleagues (1997) presented results in their study from both the standard negative binomial model with and without random effects for unobserved heterogeneity. A comparison of the log-likelihoods from the two models was used to test for the presence of significant persistent unobserved heterogeneity in the data, and a comparison of the results led the authors to conclude that there was a highly significant level of unobserved heterogeneity present in the data. The link between past and subsequent criminal activity was
ascertained through the parameter estimate for the variable indicating whether the individual had been arrested in the previous year. The parameter estimate for the ‘state dependence’ variable was 0.631 and highly significant (t-value = 8.82) in the negative binomial model that only controlled for heterogeneity through the inclusion of observed (measured) covariates such as previous juvenile adjudications, race, gender, child abuse, family structure variables, and parental criminality. The parameter estimate was reduced to 0.228 in the random effects model, allowing for both measured and unmeasured heterogeneity, thus indicating a substantial reduction in the magnitude of this effect (64% reduction in absolute size) after controlling for persistent individual differences. However, it should be noted that there was still a significant and positive effect even after allowing for persistent individual differences in the proclivity to offend. Thus, these results also indicate support for the mixed position that allows for both population heterogeneity and state dependence processes as causes of criminal offending.

In the aforementioned study, Paternoster et al. (1997) further tested for differential effects of the state dependence process between the ‘life-course-persistent’ and the ‘adolescent-limited’ offenders as hypothesized by Moffitt (and by Patterson). Recall that the arguments of Moffitt (1993, 1997) led to the conclusion that the offending patterns of the life-course-persistent group should be dominated by a static, population heterogeneity process (that has run its full course by the end of childhood/beginning of adolescence), whereas the offending patterns of the adolescent-limited group should be dominated by the state dependence processes and should be relatively unaffected by variables representing individual differences.

Age at first adjudication was used here by Paternoster and his colleagues (1997) as a proxy variable representing whether the case is a life-course-persistent (high criminal propensity) or an adolescent-limited (low criminal propensity) offender. In a series of models (fourteen separate models to be exact), the authors use different cut-points for the age at which to divide the sample into the ‘low’ and ‘high’ criminal propensity groups, and then tested for differential state dependence effects on the basis of models run on the ‘low’ and ‘high’ criminal propensity groups separately.19

---

19 Specifically, they tested whether the difference in the effect of the indicator of arrest in the prior period was statistically different from zero between the two groups.
The authors found age 15 to be a cut-point that generated different (arguably minor) state dependence effects between these two groups. The authors also found, however, that any estimated differential effects were highly sensitive to the age (at first adjudication) used to divide the sample into the two groups. As noted by Paternoster et al. (1997: 261), ‘this result [the age 15 cut-point] strikes us as being consistent with the predictions offered by developmental theorists, but the lack of robustness in this result due to slight variations in the early/late onset sample division scheme leaves us with some question as to whether the result is artifactual.’

In a different study using the panel self-report data from the National Youth Survey, Paternoster and Brame (1997) investigated the relationship between past and subsequent criminal activity among a sample of more ‘conventional’ youths than those studied in the earlier Paternoster et al. (1997) study. This study used the 479 respondents who were aged 11 or 12 in the first wave of the NYS study and followed them over the next four waves of the study (to age 15 or 16). Paternoster and Brame estimated random effects probit and negative binomial models on the binary and count variables, respectively, reflecting self-reported delinquent activity. The delinquent acts used in constructing the measures were theft exceeding $5, motor vehicle theft, aggravated assault, sexual assault, gang fights, robbery, and breaking and entering (e.g. burglary). Paternoster and Brame were interested in the effects of two state dependence variables: a binary variable indicating delinquent behavior in the prior period (wave) and exposure to delinquent peers (proportion of friends who engage in delinquent acts).

The results of their analyses indicated that both prior delinquent activity and exposure to delinquent peers were positively and significantly related to current criminal activity. Controlling for persistent unobserved heterogeneity, which was found to be highly significant, the parameter estimate of the indicator of criminal activity in the prior wave was 0.86 in the probit model (participation model) and 0.984 in the negative binomial model (frequency model). The estimate from the logit model was very similar in magnitude to the estimate of Nagin and Paternoster (1991) in their analysis of the offending patterns of another ‘conventional’ sample. Similar to the Paternoster et al. (1997) study, Paternoster and Brame also test as to whether separate, distinct models (as hypothesized in Moffitt’s theory) are necessary for describing the
offending patterns of the life-course-persistent (‘early starter’) and adolescent-limited (‘late starter’) offenders. After dividing the sample into two groups on the basis of their ‘offending propensity’ at age 12, separate models were estimated on the two groups. The findings of these analyses also indicate (like the Paternoster et al. [1997] study) that there is no strong ‘statistical evidence that these dynamic variables [prior offending and delinquent peers] exert different effects within groups of youngsters defined on the basis of their offending propensity at ages of 11 or 12 years and followed well into adolescence’ (Paternoster and Brame 1997: 74).

More recently, in a discussion and comparison of three different methodological approaches to studying the relationship between past and subsequent criminal activity, Bushway, Brame, and Paternoster (1999) have presented an empirical application of the three different statistical approaches (random effects, semiparametric random effects, and fixed effects models) with the 1958 Philadelphia Birth Cohort data.\textsuperscript{20} Using the 13,160 males in the birth cohort and seven periods of data covering ages 6–26 (each period covered three ‘age years’), Bushway, Brame, and Paternoster (1999) apply a parametric random effects probit model, a semi-parametric random effects probit model (like that used by Nagin and Land), and a conditional fixed effects logit model to a binary indicator of police contact during each period.\textsuperscript{21} This was the first published presentation and application of the conditional fixed effects model as a potential methodological approach to study the

\textsuperscript{20} Recall that Land, McCall, and Nagin (1996) also modeled the longitudinal offending patterns of the 1958 Philadelphia Birth Cohort. Land, McCall, and Nagin (1996) did include a state dependence variable (lagged indicator of conviction in the prior period) in their specification, but they did so within the ‘intermittency’ portion of their model that only included controls for observed heterogeneity (through the inclusion of observed variables) rather than unobserved heterogeneity. Nonetheless, the parameter estimate for the state dependence variable (in the intermittency part of their model) was 0.907 and highly significant.

\textsuperscript{21} Conceptually, the conditional fixed effects logit model controls for persistent (time-stable) unobserved heterogeneity through the inclusion of a separate ‘intercept’ or constant for each individual, although for numerical reasons these ‘intercepts’ are ‘conditioned’ out of the likelihood function during estimation. In other words, this estimator makes no assumption about the mixing distribution. However, a significant limitation of the model is that it does not permit the use of exogenous variables including age or trend variables. As noted by Bushway, Brame, and Paternoster (1999), the strong age or ‘trend’ effects associated with criminal activity makes this a serious limitation of this model (see also Maddala 1987; Hamerle and Ronning 1995).
processes of continuity and change in criminal offending over time. Similar to previous studies, the state dependence effect was modeled through the inclusion of a binary variable indicating police contact in the prior time period.

Results from two random effects probit models were presented first. The first model did not control for age (‘trend’) effects, and resulted in a parameter estimate for the state dependence variable (police contact in prior time period) of 1.052, which was highly significant (t-value = 48.13). The second model, which did control for time trend effects, produced a numerical estimate of 0.611 (t-value = 25.23) for the lagged police contact variable, indicating that a portion of the state dependence effect was partially the result of general temporal shifts in the probability of police contact. Both of these random effect models produced highly significant estimates of persistent unobserved heterogeneity (rho = 0.120 and 0.331 in the first and second model, respectively).

Next, semiparametric probit models with two (no time trend controls) and three (time trend controls) points-of-support were applied to the data. In the model with no time trend controls, the state dependence effect estimate was 1.035 (t-value = 49.25), while in the model with time trend controls the estimate was 0.608 (t-value = 23.72). It is interesting to note the nearly identical estimates of the state dependence effects from both the parametric and the semiparametric formulations of the probit model.

The results of the conditional fixed effects logit model, which specifically limits the independent variables in the model to the lagged dependent variable only, estimated the state dependence effect to be 1.591. Finally, after translating the ‘logit’ coefficient into ‘probit units’ by dividing the estimate by 1.6, the estimate was essentially identical (0.994) to the estimates obtained from the parametric and semiparametric models with no trend controls.

**Discussion and Hypotheses Related to the Past and Subsequent Crime Relationship**

There are two clear themes in the past decade of research on the relationship between past and subsequent criminal activity. First, there is unquestionably strong evidence of a significant amount of population heterogeneity in the propensity to commit criminal acts. Population heterogeneity was found to be significant in the
sample of ‘conventional’ respondents (Paternoster and Brame 1997), samples that over-represent individuals from an urban area (Bushway, Brame, and Paternoster 1999) and also in a predominantly working-class area (Land and Nagin 1996; Nagin and Farrington 1992a), as well as in samples consisting of ‘high-risk’ youthful offenders (Paternoster et al. 1997; Sampson and Laub 1993). Only one study (Nagin and Paternoster 1991) failed to uncover a statistically significant amount of unobserved heterogeneity in their sample. From the state dependence perspective, prior involvement in crime exerts a real (causal) effect on subsequent criminality, possibly through its attenuating effects on the social bond, and the constraints it places on future legitimate opportunities (Sampson and Laub 1993), and/or its disruptive effects on ‘local life circumstances’ (Horney, Osgood, and Marshall 1995). Despite the findings of significant population heterogeneity in offending patterns over time, Nagin and Paternoster (2000) have noted that the challenging assertions of Gottfredson and Hirschi (1990) are critical to the field of criminology (both theoretically and empirically) because they have forced the discipline to acknowledge the importance of early life events, especially those within the family, and to consider how those events may have enduring consequences for individual criminal behavior over time.²² In other words, the controversial and provocative theoretical arguments of Gottfredson and Hirschi (1990) have helped move the theoretical and empirical ‘lenses’ of criminologists away from being obstinately fixated on the adolescent years only.

A second theme in the literature of the relationship between past and subsequent criminal activity stresses the importance of state dependence processes in the lives of offenders. All of the studies reviewed above report statistically significant evidence in support of the state dependence position. That is, controlling for unobserved heterogeneity in the propensity to offend, previous criminal activity was still positively and significantly related to the probability or frequency of current offending. Thus, despite individual differences in the propensity to offend, changes in the lives of offenders appear to have important influences on subsequent criminal activity. Furthermore, these changes appear to be beyond the explanation of

²² Cohen (1987) has made similar arguments about the theoretical importance of the Wilson and Herrnstein (1985) population heterogeneity theory.
a pure population heterogeneity argument. From the state dependence perspective, prior involvement in crime exerts a real (causal) effect on subsequent criminality though its attenuating effects on the social bond, and the constraints it places on future legitimate opportunities (Sampson and Laub 1993), and/or its disruptive effects on ‘local life circumstances’ (Horney, Osgood, and Marshall 1995). The empirical evidence thus indicates that whatever one’s initial risk of crime, ‘things can get better and they can get worse’ (Nagin and Paternoster 2000: 137). Nagin and Paternoster note that Sampson and Laub’s theory has been important to the field of criminology because it brought the relevance of state dependence processes back into the view of criminologists after a period of time when the trend in criminology was to ‘push the causes of crime further back in the life course’ (Grasmick et al. 1993: 5). Sampson and Laub thus reminded the discipline that events transpiring after adolescence have potentially serious and important consequences for maintaining or changing previous behavior patterns.

Clearly, however, a summary that perhaps best characterizes the current research to date on this topic is that both causal processes appear to be influencing participation in criminal activity; that is, the evidence supports the ‘mixed’ model where state dependence and population heterogeneity processes are both seen as necessary to explain both continuity and change in criminal behavioral patterns over time. In our judgement, the best example of the ‘mixed’ position is found in the two studies that compare the magnitude of the state dependence effects both prior to and after controlling for individual differences in the propensity to offend. The studies we speak of are the ones conducted by Nagin and Farrington (1992a) and Paternoster et al. (1997). In the standard probit models (without a correction for unobserved individual differences), the magnitude of the estimates prior to and after controlling for individual differences in the propensity to offend are 1.16 and 0.631 for the Nagin and Farrington (1992a) and Paternoster et al. (1997) studies, respectively; whereas in the parametric random effects models, the corresponding parameter estimates are 0.446 and 0.228. Yet, even in the face of a roughly 63 per cent reduction in the size of the parameter estimates (after controlling for unmeasured individual differences), the state dependence variables in both studies still remained positive, significant, and substantively meaningful. Thus, it appears that just as the pure
state dependence perspective must concede that a significant amount of the link between past and subsequent criminal behavior is due to persistent individual differences in the proclivity to offend, the pure population heterogeneity perspective must concede that prior individual differences cannot explain all of the association between criminal activity at different points in time.

Recently, however, studies by Brame, Bushway, and Paternoster (1999) and Bushway, Brame, and Paternoster (1999) have each raised concerns about the validity of these important substantive conclusions due to possible artifactual flaws in prior research designs. These potential flaws concern the two main assumptions of the parametric random effects models regarding: (1) the distribution of the unobserved heterogeneity (i.e. the mixing distribution) and (2) the initial conditions assumption.

First, the possible problem with the assumption concerning the mixing distribution is that the correct statistical inferences concerning (dynamic) state dependence variables require that the mixing distribution be correctly specified (Bushway, Brame, and Paternoster 1999). As it now stands, currently, there is no agreed upon distribution assumed to correctly capture the distribution of criminal propensity in the population (Land and Nagin 1996; Nagin 2004), and, further, the nature of the distribution may be very different in ‘low-risk’ samples compared to that in ‘high-risk’ samples. It was such uncertainty regarding the actual mixing distribution in the population that led to Heckman and Singer’s (1984) ‘point-of-support’ approximation (subsequently generalized by Nagin and Land to mixtures of Poisson models for event count data) whereby the continuous distribution, whatever its shape, is approximated by a discrete, multinomial distribution. The failure to correctly specify the unobserved heterogeneity distribution may result in both biased estimates and/or inflated significance tests (Bushway, Brame, and Paternoster 1999; Heckman and Singer 1984; Land, McCall, and Nagin 1996). For example, simulations by Bushway, Brame, and Paternoster (1999) showed that, when the actual distribution of unobserved heterogeneity becomes more skewed relative to the assumed normal distribution, the bias in the state dependence parameters becomes larger, thereby unjustly favoring the state dependence explanation.

Second, the initial conditions assumption requires that the behavioral process under study (e.g. criminal offending here) be
observed at the initial start of the process (Hsiao 1986). Under this assumption, the lagged value will be zero for all cases during the first period under study (precisely because the process has not started). This condition ensures the lack of correlation between the lagged value in the first period and the time-stable (individual-specific) component of the error term in the model. Hsiao (1986) shows that it is this initial conditions process that allows the error term to fully incorporate heterogeneity in individual differences, and if one can meet this assumption, then the effect of the lagged value on the current value will be consistent, even if the lagged outcome value in subsequent periods is correlated with the persistent unobserved heterogeneity (Brame, Bushway, and Paternoster 1999).

As shown by Heckman (1981c) and Hsiao (1986), then, the main problem with violating the initial conditions assumption is that the parameter estimate for the lagged values of the outcome variable will be upwardly biased (i.e. favoring the state dependence perspective). The simulations of Brame, Bushway, and Paternoster (1999) provide further evidence that a failure to meet the initial conditions assumption upwardly biases the estimate of the lagged value (i.e. the state dependence effect). As Brame, Bushway, and Paternoster (1999: 612) note, ‘the failure to observe initial conditions guarantees that the parameter estimates from the random-effects model will be biased and inconsistent.’ The upwardly biased estimate is a direct consequence of the confounding of prior offending with the unobserved heterogeneity, whereby the parameter estimate for the lagged value absorbs some of the variation that should be rightly attributed to the time-invariant (individual-specific) component of the error term that represents population heterogeneity (Heckman 1981c).

Brame, Bushway, and Paternoster (1999) re-analyzed the data from the Paternoster and Brame (1997) study to see if the violation of the initial conditions in the data led to any erroneous conclusions regarding the impact of the state dependence variables. Using Heckman’s (1981c) approximation method, developed to correct for violations of the initial conditions assumption, they found a further reduction in the importance of the state dependence variables (lessening the impact of the delinquent peer exposure variable and a complete reduction to non-significance of the prior offending variable) after the application of this correction. The authors thus concluded that ‘reported coefficient estimates for dynamic factors
could be biased because of problems with initial conditions’ (Brame, Bushway, and Paternoster 1999: 636).

As a result of such analyses, a general doubt lingers in the field regarding the robustness and validity of the findings of previous studies: ‘in the absence of clear knowledge about fidelity to model assumptions, researchers should adopt a healthy amount of skepticism in their observed findings’ (Nagin and Paternoster 2000: 140). Supporting this notion (in a critical essay on the superfluous treatment of the assumptions of statistical models), Maltz (1994) has persuasively warned that criminologists must devote more attention to checking the assumptions of the statistical models they apply to crime data or risk possibly generating publishable yet erroneous/invalid results. Consider, for example, the two following points regarding the seven primary studies reviewed above. First, five of these studies rely entirely on the parametric random effects model and make no attempt (for obvious reasons of both data and software limitations) to check the robustness of their findings to the assumed distribution of the unobserved heterogeneity in their data. The two notable exceptions that do take advantage of the semiparametric formulation of the random effects model are the studies of Land and Nagin (1996) and Bushway, Brame, and Paternoster (1999). While the Nagin and Farrington (1992a) study did make use of the same data set as did Land and Nagin (1996), the state dependence variable relating past to current participation in criminal activity was specified differently between the two studies. The results of the analyses presented by Bushway, Brame, and Paternoster (1999) yielded virtually identical numerical and substantive results in both the parametric and semiparametric formulations of their statistical models. To date, then, the only study that has calculated the state dependence effect (with the same data and the same exact model specification) with both the parametric and semiparametric random effects models yielded identical substantive results for both models. As Bushway, Brame, and Paternoster (1999: 53) state, however, ‘since there is no reason to believe, a priori, that the results of our substantive analyses are generalizable beyond the specific data set that we used, we think that multiple-method strategies for investigating questions such as the one addressed here…are necessary.’ Thus, the degree to which the assumptions of the statistical models yield any substantive differences in the conclusions requires testing with other data sets such as those to be used in this study.
Second, only two of the six studies that employ the parametric random effects model use data that unquestionably do not violate the initial conditions assumption. These two studies are those of Nagin and Farrington (1992a) and Bushway, Brame, and Paternoster (1999). In the other four studies, the offending process had already been initiated at the point in time that each study began their panel, and thus there is a possibility that the estimates of the variables representing the association between past and subsequent criminal activity would be upwardly biased.

Indeed, the two studies using samples of high-risk respondents (Paternoster et al. 1997; Sampson and Laub 1993) relied entirely on the parametric random effects model and began their panel studies at a point in time when all of the respondents had already begun their criminal offending. Given the findings of Brame, Bushway, and Paternoster (1999) and Bushway, Brame, and Paternoster (1999) regarding the consequences of violating, these two aforementioned studies are suspect. This point is fundamentally critical because as Nagin and Farrington (1992a), Paternoster and Brame (1997), and Nagin and Paternoster (2000) have all hypothesized, the importance of state dependence variables may depend on the nature of the sample employed in one’s study. Again, this assertion is based on the fact that studies showing stronger support for the state dependence perspective tend to employ ‘low-risk’ samples (and self-report data), whereas studies showing stronger support for the population heterogeneity perspective tend to employ ‘higher-risk’ samples and to use official statistics such as arrest or conviction data.23 Paternoster and Brame (1997) speculated that such findings are consistent with the theoretical propositions of Moffitt’s (1993) dual taxonomy theory because samples consisting of ‘higher-risk’ cases should contain a larger percentage of ‘life-course-persistent’ offenders (whose behavior is governed by static processes), whereas ‘adolescent-limited’ offenders (whose behavior is governed by dynamic processes) should constitute the majority of the respondents in ‘low-risk’ samples (see also Cernkovich and Giordano 2001).

In sum, there currently is some ambiguity about the state of the extant empirical evidence regarding the relationship between

---

23 Again, it is important to bear in mind that while nearly all studies find evidence in support of both positions; it is the degree of support for each position that this issue concerns.
past and subsequent criminal activity, especially with respect to how robust the findings are with respect to assumptions of the statistical models employed and how important state dependence processes are in the population of ‘high-risk’ offenders. We concur with Nagin and Paternoster (2000: 131) that ‘only by examining the sensitivity or robustness of research findings with different statistical models and different data can the field hope to come to an understanding about the tenability of population heterogeneity and state dependence.’ Cernkovich and Giordano (2001) also have recently commented that there simply is ‘scant evidence’ regarding the empirical importance of these two explanations (state dependence and population heterogeneity) for continuity and discontinuity of criminal offending patterns across the life course, especially in both the serious offender population and in data sets that include ages beyond adolescence (see also Farrington 2003).

In direct response to these calls for further investigations of this key theoretical issue, our study will test the following four hypotheses concerning the relationship between past and subsequent criminal offending behavior:

**H$_5$:** There will be a statistically significant positive association between past and subsequent offending behavior.

**H$_6$:** After controlling for persistent individual differences in criminal propensity, the association between past and subsequent offending will be reduced to a non-significant level.

**H$_7$:** After controlling for persistent individual differences in criminal propensity, the association between past and subsequent offending behavior will be reduced in magnitude but will still be positive and statistically significant.

**H$_8$:** The association between past and subsequent offending behavior will be non-significant for the ‘life-course-persistent’ (high criminal propensity) group(s), while the effect should be substantial and significant for the ‘adolescent-peaked’ group.

The key hypotheses for the three theories discussed in this study are **H$_6$, H$_7$, and H$_8$.** Evidence supporting **H$_6$** would lend credence to the theory of Gottfredson and Hirschi’s self-control theory. Evidence supporting **H$_7$** would be consistent with the predictions of Sampson and Laub’s age-graded social control theory, and evidence supporting **H$_8$** would appear to validate Moffitt’s dual taxonomy theory.
Results concerning these four hypotheses are presented in Chapter 8 of this study. Here we will draw and build on the multi-method approach of Bushway, Brame, and Paternoster (1999) to test $H_6$ and $H_7$. More specifically, we will use both the parametric negative binomial random effects model and the semi-parametric mixed Poisson model of Nagin and Land to test these hypotheses. In addition, we will also employ standard negative binomial models with a set of binary variables that indicate latent class membership (from the latent class results presented in Chapter 7) to more definitively test the robustness of the presence (or absence) of state dependence effects in a longitudinal panel analysis of the offending patterns for three samples of California Youth Authority parolees.

To test the last hypothesis ($H_8$) separate models will be estimated on offenders assigned to a given latent class. It should be noted that $H_8$ is a conditional hypothesis that requires the identification of an ‘adolescent-peaked’ group in the data sets. To date, tests concerning $H_8$ have been accomplished after dividing the samples into two groups on more arbitrary grounds (e.g. age of onset) rather than by calculating the effects within a group shown to actually offend in an ‘adolescent-limited’ fashion. Before concluding the present chapter and moving on to Chapter 4, where we briefly describe the California Youth Authority, we discuss the potential contributions this study can make to the discipline of criminology.

**Contributions to the Discipline**

The proposed research conducted herein will attempt to make several contributions to the discipline of criminology. These contributions include a general accretion of knowledge to the study of the continuity and discontinuity of criminal offending patterns over the life course of chronic serious youthful offenders, as well as specific contributions that advance our current knowledge concerning both the relationships between age and crime and between past and subsequent criminal activity. A major potential contribution of this study centers on the nature of the samples employed in our analyses: the fact that we have acquired three separate samples from three different time periods, that the subjects in our study are followed over a longer period of time than most of the previous studies, and that the sample sizes employed here are larger than
those used in previous studies makes our study unique. Youthful offenders who commit the most serious crimes at a disproportionately high rate have to date been largely unavailable to social scientists (see, e.g., Cernkovich, Giordano, and Pugh 1985; Cernkovich and Giordano 2001; Laub and Sampson 2001), and therefore a detailed empirical analysis of their longitudinal criminal offending patterns will provide social scientists and policy makers with a more accurate characterization and deeper understanding of the longitudinal patterns of criminal activity across the life course of this select group of offenders.

To be more specific, we note that to date much of our knowledge concerning the serious, persistent young offender has been derived through analyses of the most frequent offenders (usually referred to as the chronic offenders) in birth cohort studies such as the 1945 and 1958 Philadelphia birth cohorts and with general population samples. The major finding of these Philadelphia birth cohort studies was that roughly 6 to 7 per cent of the individuals in the cohorts were responsible for more than half of all of the officially recorded police contacts reported for these cohorts (see, e.g., Tracy, Wolfgang, and Figlio 1985; Wolfgang, Figlio, and Sellin 1972). For example, Wolfgang and his colleagues report in the 1972 study that these young ‘chronic criminals’ were responsible for committing 63 per cent of all known Index crimes committed by the birth cohort members (as measured through police contacts)—including 82 per cent of all robberies, 73 per cent of all rapes, 69 per cent of all aggravated assaults, and 71 per cent of the murders.

However, several re-analyses of the 1945 birth cohort data question exactly how chronic or serious most of the chronic offenders in these types of samples (e.g. birth cohort studies and/or national probability samples) really are. For example, analyses by Bernard and Ritti (1991) indicated that only 35 per cent of all police contacts in the Philadelphia birth cohort ever resulted in an actual

24 The bulk of the analyses in this study involve the longitudinal offending patterns of 4,866 sample subjects (n = 1,989 in the 1981–2 sample, n = 1,443 in the 1986–7 sample, and n = 1,434 in the 1991–2 sample).

25 It was findings such as these that initially aroused interest in the application of selective incapacitation policies. Of course, selective incapacitation policies will only produce their desired effect if the high-rate offenders commit crimes at a relatively constant, stable rate across the age distribution (see Ezell and Cohen 1997; Gottfredson and Hirschi 1986; Haapanen 1990).
formal arrest, and that an astounding 31 per cent of the 627 ‘chronic’ offenders (those with five or more police contacts) identified in the 1972 study were either never arrested \( n = 48 \) or arrested only one time \( n = 145 \) in their entire juvenile criminal career. Similarly, analyses by Weitekamp et al. (1995) showed that 73 per cent of the aggravated assaults in the 1945 Philadelphia birth cohort were committed by 32 of the 627 chronic offenders and 71 per cent of the homicides were committed by only 10 of the chronic offenders. In addition, studies of criminal offending that employ the use of general population samples generally contain too few serious youthful offenders (because of their low base rate in the population) to allow for reliable descriptions and investigations of their offending behaviors over time (see, e.g., Cernkovich, Giordano, and Pugh 1985; Loeber and Farrington 1998). For comparative purposes, we inform our readers that it has been estimated that only 1 out of every 1,000 police contacts will result in a single case being committed to the California Youth Authority (Legislative Analysts Office 1995).

Second, the length of time that the cases in our three samples will be longitudinally tracked allows for rigorous testing of the extent and nature of the patterns of continuity among this population. Standard (between-individual) empirical assessments of continuity in offending behavior often use cross-sectional data and/or short-term panel data that preclude most studies from addressing the main questions to be investigated in our study (as they generally require the use of extensive longitudinal data):

Most criminological research consists of cross-sectional ‘snap-shots’ or short-term panel studies of crime over the full life span. As a consequence, relatively little is known about desistence and, for that matter, the processes of persistent criminal behavior through the life course. Indeed, the characteristics that distinguish persistence in a life of crime from desistence within any group of high-risk offenders are generally unknown.

(Laub and Sampson 2001: 1)

Our study will attempt to addresses this limitation in the empirical literature by following three samples of chronic serious offenders from the date of their first recorded arrest up until 30 June 2000.\(^{26}\)

\(^{26}\) For the three release samples combined, the earliest year of birth was 1956 and the latest was 1978. The 25th percentile for the year of birth was 1963, the median was 1967, and the 75th percentile was 1971.
As we demonstrate later, the follow-up period for our data is considerably longer than the previous studies reviewed above.

Laub and Sampson (2001) also point out that our knowledge of the long-term offending patterns among serious offenders has been hampered, not only by a lack of studies that longitudinally follow this group for extended periods of time, but also because of the disjunction between the juvenile and adult record systems. Crime data often suffer from the division of juvenile and adult criminal record-keeping systems, meaning that there is often a dearth of data bridging the juvenile and adult years in longitudinal studies and that many data sets are often blind to criminal activity on the other side of the juvenile–adult age boundary (Blumstein et al. 1986; Cernkovich and Giordano 2001; Laub and Sampson 2001). The research undertaken in our study is unique because it entails following three samples of youthful offenders from the date of their initial arrest, through the year(s) they were incarcerated in the California Youth Authority, and then into adulthood, and well into adulthood for some individuals (age 43 was the oldest age). The first sample in our study was released on parole in fiscal year 1981–2 and was followed up through their late thirties to early forties (depending on age at time of release from the CYA). The second sample was released in fiscal year 1986–7 and followed up through their early to mid-thirties. Finally, the third sample was released in fiscal year 1991–2 and followed up through their mid- to late twenties.\(^{27}\) We believe that these data sets comprise one of the most comprehensive sets of longitudinal data on the serious youthful offender population that has yet been gathered. In sum, the frequency and seriousness of criminal activity displayed among our samples, the length of the follow-up period employed here, and the sophistication of the statistical models employed in our study should allow for a rigorous examination of criminal offending patterns over the life course with respect to the critical substantive and methodological issues identified in this chapter.

The empirical fact is that we currently know very little about the offending patterns of very serious youthful offenders; we simply lack evidence regarding fundamental questions concerning the criminal offending patterns of these individuals. In a recent book

\(^{27}\) For the 1981–2, 1986–7, and 1991–2 samples, the average ages at the end of follow-up period (30 June 2000) were 37, 33, and 27, respectively.
surveying the current state of empirical knowledge on serious and violent juvenile offenders, the editors concluded with a section entitled, ‘Developing a Research Agenda.’ Here, they noted that there are currently many ‘gaps’ in the knowledge concerning the nature and development of the longitudinal criminal offending patterns of serious and violent juvenile offenders, including the validity of offender typologies (e.g. ‘life-course-persistent’ and ‘adolescent-limited’ offenders) in this population, the nature and extent of the adult criminal offending patterns and adult life experiences of this group of offenders, and which covariates predict a continued persistence in offending within this group (Loeber and Farrington 1998). Similarly, Laub and Sampson (2001: 10) call for a theoretical and empirical focus on the patterns of continuity (persistence) and discontinuity (desistence) among samples of persistent and serious offenders, noting that ‘criminologists should not spend much time or energy studying termination and desistence for low-rate offenders’ because such offending is normative during adolescence. The analyses in this study hence should provide needed evidence concerning the nature and extent of criminal offending across the life courses of the most serious youthful offenders.

In sum, we currently have very limited information concerning the actual shape of the aggregate age–crime curve within the serious offender population, on what ‘latent classes’ of offenders are present in such a population, and if the identified latent classes are resilient across time. As Cernkovich and Giordano (2001: 405–6) note:

if we accept the premise that there is a small group of offenders who do not begin and age out of crime in the same fashion as most offenders, then it is important that researchers examine in detail the extent to which the stability–change paradox is a function of the existence of two distinct populations of offenders . . . however this issue has not been systematically examined, in part because of the relatively scant (though increasing) research focusing on serious chronic offenders . . . it is essential that the research agenda be expanded to include an even greater focus on this group and . . . that it include attention to patterns of antisocial activity prior to and beyond the adolescent years.

Furthermore, even though there is a group of offenders who have been semantically labeled ‘life-course-persistent’, to date there is no convincing empirical evidence to prove that this group of offenders is deserving of such a demonstrative label. A key question
yet to be answered is: how persistent are the ‘life-course-persistent’ offenders? Due to the highly selective nature of who gets committed to the California Youth Authority, samples of offenders released from the CYA have a unique potential to address this question. The data we describe in Chapter 6 aim to address these specific limitations by applying Nagin and Land’s finite mixture model to the three large samples of serious youthful offenders that are followed for an extensive period of time.

Our review of the prior studies of the relationship between past and subsequent criminal activity also made it clear that there is a critical need for the continued examination of this topic. As indicated above, previous studies have consistently found evidence in support of the mixed-model position; however, questions remain regarding the authenticity of the state dependence effects uncovered in the prior research, especially within high-risk samples, due to the possible methodological consequences of violating the assumptions of the parametric random effects models. The results to be presented in Chapter 8 of this study should contribute to the extant literature on this topic by examining the relationship between past and subsequent criminal offending.

Before getting to the methods and the data analysis chapters in this study, however, our attention in the next chapter is first focused on the California Youth Authority, the institution from which our samples have been released on parole.