There is a long-standing debate in criminology about the relative impact of static versus dynamic factors on criminal behavior. Researchers interested in estimating the impact of dynamic factors like prior offending or association with delinquent peers on criminal offending must control for static factors like intelligence, family background, or self-control, which could plausibly be correlated with criminal offending and the dynamic factor itself. Unfortunately, as a practical matter, it is not possible to observe all of these static factors. Statisticians and econometricians have shown that it is possible to identify the collective effect of static factors even though they cannot be observed. To achieve this objective, however, it is necessary to account for stable, unobserved individual characteristics through the use of "fixed-effect" or "random-effect" estimation. Criminologists often use random-effect estimators in these situations. We describe some of the assumptions that are necessary to develop valid inferences when time-varying covariates are used. Then, we use simulation evidence and an empirical application to show that bias can result when they are violated.

THE ISSUE

During the past two decades, criminologists have increasingly used longitudinal research designs to investigate how criminal offending develops and changes over the life span (see e.g., Blumstein et al., 1986; Elliott et al., 1985; Farrington, 1986; Greenberg, 1979; Loeber and LeBlanc, 1990; Saltzman et al., 1982; Sampson and Laub, 1993). Several empirical results have emerged from these efforts. For example, past and future criminal behavior are strongly associated with each other (Nagin and Farrington, 1992a, 1992b; Nagin and Paternoster, 1991; Sampson and Laub, 1993).
Also, delinquent peer exposure is associated with involvement in criminal activity while marital and occupational attachments seem to be associated with desistance from criminal activity (Elliott et al., 1989; Laub et al., 1998; Sampson and Laub, 1993). No one disputes the fact that it is potentially useful to know that these correlations exist. Still, whether they are of any etiological importance is quite another matter. It is on this issue that criminologists leave the realm of consensus and enter a more ambiguous realm, where different criminologists have very different views.

In this article, we attempt to bring some of the sources of this dissensus out into the open. In so doing, we consider some of the available strategies for analyzing longitudinal records of criminal behavior. One such strategy, to which we devote the bulk of our discussion, involves the use of a collection of tools that we generically describe as random-effects models. Random-effects modeling strategies are receiving increased attention in contemporary criminology and hold the potential for significantly advancing knowledge about the processes that generate criminal behavior at different points in the life span. In this article, we explore some of the properties of these models and the assumptions upon which their validity depends. We demonstrate the potential effect of particular violations of the assumptions on the validity of inferences about individual-level crime-generating processes. Finally, we illustrate the practical importance of these assumptions by reexamining some results previously reported in this journal.

THEORETICAL BACKGROUND

The strong positive association between past and future individual criminal activity is one of the most agreed, yet least well understood, facts about law-breaking behavior. Individuals who have offended in the past are most likely to offend in the future. There is little doubt or ambiguity about the validity of this claim. Still, it is not clear why this association exists. As Nagin and Paternoster (1991) observed nearly a decade ago, a number of very different explanations are all potentially capable of explaining the relationship between past and future criminal activity. For example, some theorists have argued that variation in criminal offending can be explained by variation in a time-stable predisposition to commit crimes (see e.g., Gottfredson and Hirschi, 1990; Wilson and Herrnstein, 1985). According to this point of view, differences between individuals in this propensity to offend are established early in life. Once established, these differences tend to endure. To these theorists, the relationship between past and future criminal activity is not causal. Rather, they contend, it can be explained entirely by enduring individual differences in the proclivity to offend (Nagin and Paternoster, 1991).
Similarly, such theories also discount the etiological importance of many well-documented *time-varying* or dynamic correlates of criminal offending, such as school performance, delinquent peer exposure, and marital and occupational attachments. According to these *stable-individual-difference* theories, enduring dispositions to offend are synonymous with enduring dispositions to select delinquent peers for friends. They are synonymous with enduring dispositions to perform poorly in school and in the workplace. They are synonymous with the inability to enter into warm, mutually satisfying marital relationships. The list could continue. In sum, stable-individual-difference theories maintain that the associations between well-documented crime correlates and criminal activity are spurious once relevant stable individual differences have solidified. For convenience, we will refer to criminological theories that implicate only stable individual differences as *static* explanations.

A number of contemporary criminological theorists, however, have adopted a more *dynamic* position. Research that investigates the validity of these dynamic theories will, of necessity, be longitudinal in nature. According to dynamic theorists, the temporal development of criminal activity cannot be understood by appealing only to the continuing reverberation of stable individual differences. And, it cannot be investigated by studying individuals at one point in time. Instead, dynamic theorists predict that life experiences (including criminal activity itself), changing social networks, and evolving perceptions will also explain why some people offend at particular time periods while others do not—even after the influence of stable individual differences has been taken into account (Horney et al., 1995; Laub et al., 1998; Loeber and LeBlanc, 1990; Nagin and Paternoster, 1994; Sampson and Laub, 1993).

This latter point is especially important because there seems to be widespread agreement among criminologists that stable individual differences must be taken into account when estimating the effects of variables whose values are subject to change over time (see e.g., Nagin and Farrington, 1992a, 1992b). Numerous researchers have noted that a variety of problem behaviors, such as drug use, fighting, unemployment, marital instability, and crime, tend to cluster within the same individuals (see e.g., Donovan and Jessar, 1985; Laub et al., 1998 Nagin et al., 1995; Osgood et al., 1988; Robins 1966, 1978; Rowe and Flannery, 1994). While it may be the case that some problem behaviors causally increase the risk of involvement in other problem behaviors, an equally compelling rival hypothesis is that a collection of individual characteristics causally increases the risk of involvement in many different kinds of problem behavior. And, clearly, both processes may be operating at the same time.

Therefore, it is well settled that some attempt must be made to control for stable individual differences when estimating the effects of dynamic
factors on criminal behavior. In practice, however, not all of the relevant stable individual differences are observed. That they are not observed does not, of course, make them any less real, and therefore, it is also well settled that attempts must be made to control for stable but unobserved individual differences.

In sum, dynamic theorists would adopt the position that the resolution of the relative validity of static and dynamic explanations of criminal offending requires two things: (1) measures of individuals over time and (2) the ability to assess the impact of dynamic factors (e.g., job stability, marital satisfaction, and so on) on criminal offending net of stable individual differences. Sometimes, these individual differences are observed because there are available measures of things that capture criminal propensity (e.g., intelligence, parental criminality, impulsivity). More often, however, such measures are not available or are incomplete, and researchers must adopt some method for holding them constant. As we discuss in the next section, static theorists are generally of a different mind regarding the resolution of the debate.

METHODOLOGICAL IMPLICATIONS

The theoretical dispute about the etiological import of static and dynamic crime-producing factors has important methodological implications. In stark contrast to the position of dynamic theorists, the clear methodological implication of a static explanation is the adequacy of cross-sectional studies of crime and the corresponding irrelevance of longitudinal or panel studies. Gottfredson and Hirschi (1990:232), for example, argue that if differences in crime are due to stable individual differences in crime-proneness, longitudinal designs are simply unnecessary for drawing accurate inferences about processes that generate criminal behavior over the life span: "If there is continuity over the life course in criminal activity (or its absence), it is unnecessary to follow people over time." Their argument can be summarized as follows: (1) issues of cause and effect are not easily resolved by nonexperimental longitudinal research designs; (2) current research strategies with panel data amount to little more than repeated cross-sectional analysis; and (3) the task of sorting out the validity of static and dynamic explanations of criminal behavior may best be resolved by experimental research.

In our view, Gottfredson and Hirschi make an important point. Other things being equal, experiments clearly provide a more credible basis for making causal inferences than panel designs. However, as Gottfredson and Hirschi (1990:220) acknowledge, it is difficult to think of ways to create experiments to rigorously test key propositions of dynamic theories (or, to falsify static ones). How, for example, can one randomly assign
individuals to different kinds of marriages or different kinds of peer
groups? This leaves an unhappy choice between nonexperimental cross-
sectional designs and nonexperimental longitudinal designs. The question
for criminologists, then, is whether one choice is more unhappy than the
other. Gottfredson and Hirschi contend that the two designs are equally
incapable of identifying causal effects, and cross-sectional designs should
be preferred because of their lower cost. But, as they readily grant, this is
a minority opinion in criminology (Blumstein et al., 1986; Farrington, 1986;
Petersilia, 1980; Sampson and Laub, 1995).

A critical limitation of conventional cross-sectional analysis in distin-
guishing between static and dynamic explanations is that it does not allow
estimation of the effects of dynamic factors while controlling for stable
individual differences. In other words, cross-sectional designs cannot dis-
tinguish between the effects of variables whose values change over time
and variables whose values do not. In fact, cross-sectional designs treat
every explanatory variable as a static factor. But, when properly con-
ducted, studies of panel data have the virtue that they can distinguish
between the effects of variables whose values change over time and the
effects of variables whose values are constant over time. If panel studies
can help criminologists develop valid inferences about the effect of
dynamic factors on criminal behavior while controlling for stable between-
individual differences, the additional expense would seem to be justifiable.

If panel studies are potentially so helpful in assessing the merits of static
and dynamic explanations of criminal behavior, why do Gottfredson and
Hirschi seem to dismiss out of hand the utility of longitudinal designs? We
think their position is more complex than a blanket rejection of panel
analyses. Gottfredson and Hirschi's view has always been that the current
state of longitudinal analysis in criminology has failed to "deliver the
goods" (where the “goods” are the ability to develop causal inferences
from nonexperimental data). In our view, Gottfredson and Hirschi's con-
cerns about basing inferences on longitudinal analyses are more produc-
tively seen as a challenge to the field than as anti-longitudinal dogma. In a
commentary on the problem of developing valid causal inferences from
longitudinal data, Hirschi (1987:200) observed that “it may not be possible
to solve this problem by analysis of the data reflecting it, but it would be
possible to do a better job.” In sum, Gottfredson and Hirschi seem to
place the burden squarely on the shoulders of those who would do longitu-
dinal analysis to demonstrate that the greater expense and labor inherent
in the design are justified by the production of insights that cannot be
obtained with cross-sectional analysis.

In this article, we respond to the challenge of Gottfredson and Hirschi
by examining the ability of one class of statistical models (as they are conventionally applied to longitudinal data) to help us develop valid inferences about the effects of dynamic factors on criminal behavior. Specifically, we focus on the so-called random-effects model. Random-effects modeling strategies provide criminological researchers with a very powerful set of tools for investigating the development of criminal offending over the life span. Based on our review of the criminological as well as the relevant statistical and econometric literature, we are convinced that the validity of inferences from random-effects models depends to some extent on assumptions that are not usually critically examined. Our own analysis convinces us that if these assumptions are not met, parameter estimates obtained from random-effects modeling efforts are likely to be biased. Moreover, the direction of the bias will tend to favor a dynamic rather than a static explanation of criminal behavior—a finding that supports Gottfredson and Hirschi’s skepticism about traditional panel-data analysis. But we also show that studies of longitudinal data sets can lead to approximately valid inferences about the influence of static and dynamic factors on criminal behavior under certain conditions.

The remainder of the article is organized as follows. First, we provide a brief conceptual introduction to random-effects modeling strategies. Next, we attempt to explain intuitively the technical assumptions underlying these random-effects panel models. Then, using Monte Carlo simulation evidence, we show that parameter estimates will, in general, be biased when these assumptions are not met. Finally, we explain and implement some simple methods to reduce (but not entirely eliminate) those biases. We illustrate the application of these methods with more simulation evidence; we then revisit some analysis results that we previously published in this journal.

RANDOM-EFFECTS STATISTICAL MODELS

Researchers are often interested in estimating the effects of variables whose values vary over time on criminal behavior. In criminology, classic examples of such variables include school performance, marital status, employment, delinquent peer exposure, and prior offending experience. The problem with obtaining valid inferences in these settings is that time-stable sources of criminality may be responsible for explaining at least some of the variation in all of these variables. Under these circumstances,

1. In this section, we try to provide some intuitive discussion about basic issues involved in random-effects model estimation. More formal treatments of these models can be found in most intermediate econometrics and statistics textbooks (see e.g., Greene, 1997; for a more specialized treatment see Hsiao, 1986).
at least some of the observed association between these time-varying variables and criminal behavior is spurious. If some of the confounding time-stable sources of criminality are not observed, it becomes more difficult to obtain valid inferences about the effects of time-varying variables.

Econometricians and statisticians have developed a number of methods that lead to valid inferences about the effects of observed time-varying variables in the presence of unobserved time-stable variables that are correlated with included time-varying variables. The most prominent of these methods are fixed-effects and random-effects panel models. Fixed-effect estimation undoubtedly provides the strongest control over the confounding influences of unobserved time-stable sources of criminality. The goal of fixed-effect estimation is to estimate a term for each individual in the data that takes all of the stable features of the process under study (i.e., criminal behavior) into account. This term is called a “fixed effect” and the data set can then be characterized by its distribution of fixed effects. After the fixed effects have been taken into account, the only variation that remains is variation due to individual change over time. Since the focus of this estimator is on explaining individual changes after the influence of stable individual differences has been removed, the estimated effects of variables in fixed-effects models are free of any confounding or bias that might be induced by these stable individual differences (see Allison, 1994, for a very helpful discussion of fixed-effect estimators).

While fixed-effect estimation clearly provides the strongest control over stable individual differences, it has some limitations in criminological work. First, it is more difficult to implement fixed-effect estimation when the outcome under study is qualitative rather than quantitative. Much criminological work focuses on variables such as whether an individual offends or how many times an individual offends within some finite time period. As King (1989) points out, it is not realistic to treat these sorts of discrete outcomes as realizations of a continuous data-generating process much less a continuous and normal data-generating process. It turns out, however, that traditional fixed-effect estimators are only consistent in qualitative outcome models when the number of time periods under study becomes large. In most contemporary criminological applications, the number of time periods under study is quite small (usually less than 15 to 20 periods). Thus, traditional fixed-effect estimation is not viable in this setting (see e.g., Hsiao, 1986:159; Maddala, 1987, 1993).

Some alternative fixed-effect estimators have been developed for qualitative outcomes. These methods involve the use of conditional likelihood functions (see e.g., Hausman et al., 1984; Maddala, 1987). In the case of binary or dichotomous outcomes, however, they are quite demanding in terms of the sample size because cases with certain kinds of outcome patterns drop out of the analysis entirely (Allison, 1994). In order to use
fixed-effect estimators to obtain consistent estimates about the effects of prior behavior on future behavior, it is necessary to make some strong assumptions about the absence of a linear time trend in the outcomes over the time period under study (Corcoran and Hill, 1984; Bushway et al., 1999).

Another common complaint about fixed-effects strategies is their inability to identify the effects of individual time-stable explanatory variables. Instead, the effects of all time-stable variables are captured by the "fixed effects" and are, therefore, treated as nuisances rather than as substantively interesting parameters. While this is not necessarily a bad thing (indeed, the strength of controlling for all time-stable confounding variables will often far outweigh the weakness of not estimating the effects of time-stable variables individually), it is something that should be considered when choosing an estimator. Perhaps because of some of these reasons, fixed-effect estimators have not received much attention in longitudinal research in micro-level criminological studies. Instead, most contemporary criminological work has relied on random-effects estimation.

We define a random-effects estimator as a model that assumes that individual-specific effects are randomly drawn from some well-defined probability distribution (Hsiao, 1986:32–33). Random-effects models arise in a variety of circumstances. For example, if one estimates a structural equation model (using LISREL or generalized least squares estimators) in which a dependent variable is observed at two or more time points for each individual and one allows for equal correlation between the residuals of the outcomes, then the structural equation model is a random-effects estimator (in the sense described here). One version of the hierarchical linear modeling (HLM) approach described by Bryk and Raudenbush (1992) can also be viewed as a random-effects model; specifically, when the HLM is implemented to allow for a random intercept (with no other random coefficients), it can be viewed as a conventional random-effects estimator.

Random-effects estimators have also been implemented for qualitative
outcomes within the HLM framework and in major software packages, like LIMDEP (Greene, 1995), with which random-effects probit models for binary outcomes and random-effects negative binomial models for event-count outcomes can be estimated. In standard generalized least squares, hierarchical linear models, and random-effects probit models, the available software typically operates under the assumption that the individual-specific effects are drawn from a normal probability distribution. The random-effects negative binomial model assumes that the random-effects are drawn from a beta probability distribution. Although other probability distributions for the random effects are possible (see e.g., Longford, 1995), in practice, the software to estimate models with other probability distributions is not yet widely available.

In recent years, Nagin and his colleagues have been engaged in a program of criminological research that allows for greater flexibility in the specification of the probability distribution from which the random effects are drawn (Land and Nagin, 1996; Land et al., 1996; Nagin and Land, 1993) for models with both qualitative and quantitative outcomes. In their framework, the shape of this distribution is not specified at all. It is only assumed to be discrete in form. Recent research using Wolfgang's 1956 Philadelphia cohort has compared the results using the Nagin and Land semiparametric method to the standard random-effects method (Bushway et al., 1999). The results obtained from these two estimators are virtually identical, which indicates that for at least one prominent data set, assumptions about the probability distribution associated with the unobserved heterogeneity are either not violated or violated with little consequence. We conclude, therefore, that although concern about this assumption is legitimate, other aspects of the model are perhaps more problematic. In the next section, we describe two other key assumptions of the random-effects estimator that are often violated.

ASSUMPTIONS OF RANDOM—EFFECTS PANEL ANALYSIS

ZERO COVARIANCE BETWEEN THE ERROR TERM AND TIME-VARYING PREDICTORS

A potential issue with any given choice of model specification in the social sciences is the problem of omitted variable bias. Simply put, if some elements of the process that generates the outcome are not included in the model specification and those elements are correlated with the covariates that are included in the model, the parameter estimates associated with those covariates will be biased. We wish to emphasize that this is an important issue for cross-sectional and longitudinal analyses. Specifically, there are three different types of omitted variable bias in studies like the
kind we are discussing here. 4 One type of bias occurs when time-stable omitted variables are correlated with time-stable included variables (type 1). For example, the coefficient estimate for the relationship between sex and offending will reflect both the effect of sex (a measured time-stable factor) and unmeasured levels of impulsivity (an unobserved time-stable factor correlated with sex). A second type of bias occurs when time-stable omitted variables are correlated with time-varying included variables (type 2). An example is when unobserved time-stable variables are correlated with a measured indicator of delinquent peers. A third and final type of omitted variable bias occurs when time-varying omitted variables are correlated with time-varying included variables (type 3). An example is when some or all of the effect of an included variable, like job status, is due to its correlation with an excluded variable also related to offending, such as a growing separation from unconventional peers.

In our view, type 1 bias is not unique to panel analysis, and is going to be very difficult to resolve. 5 Type 2 bias is what we are concerned about here. Type 3 bias, which we think is more tractable than type 1, is an important problem but it is beyond the scope of this discussion. 6 While we recognize that all three constitute potential pitfalls for those conducting

---

4. We thank a reviewer for emphasizing this issue.

5. This type of bias between observed and unobserved variables occurs in all types of regression analysis, not just panel analysis. However, we believe that researchers using panel data and random-effects estimation often act as if the controls for unobserved heterogeneity eliminate the problem and yield an estimate for the included variable that gives its “true” value. This is not correct, since the random-effects model requires the assumption that there is no covariance between included and excluded variables. As noted by a reviewer, the coefficient on an included time-stable coefficient like sex will still provide an estimate of the “true” sex difference, but it concerns the unconditional relationship, not adjusted for any spurious or mediating factors. That can be useful information to know, but one must keep in mind that the results are subject to bias from the omitted variables, even in the random-effects framework.

6. A reviewer of our article mentioned two methods that may be helpful in resolving this problem. One method (originally used by Heckman) relaxes the constraint that all error terms are equally correlated with all other error terms. For example, such a general specification would allow the error terms at adjacent time periods to be more strongly correlated than the error terms at nonadjacent time periods. Another method would allow the time-trend effect to be a random rather than a fixed coefficient.

The first method allows for the possibility that there are variables whose effects endure for relatively short periods of time. The second method allows for the aggregate effect of omitted time-varying variables to vary across individuals. We think these are useful suggestions. We also think that estimators that approach this problem in different ways should be systematically compared in the future. Nevertheless, for the reasons discussed in this article, our main purpose is to deal with what we think is a more urgent issue: estimating the effects of time-varying variables when there are time-stable confounding variables.
panel analyses with life-course data, we acknowledge that the methods we have described deal only with the second type. Our interest in type 2 omitted variable bias is, of course, driven by our substantive interest in clearly delineating the influence of static and dynamic factors on criminal offending. This type of bias is particularly important because it involves a confounding of time-varying (dynamic) and time-stable (static) components. We believe that resolving this issue should constitute the first step for those interested in continuity and change in criminal offending. From our perspective, we thought that we should have a handle on the potential confounding of static and dynamic effects before we could comfortably move into a discussion of the relative impact of different dynamic factors. In a nutshell, our position is that at this current state of knowledge of the confounding issue, a concern with type 3 omitted variable bias is “putting the cart before the horse.”

To make the discussion of the type 2 bias more concrete, assume that an outcome variable, $y_{it}$, is a measure of criminal activity for each of $i = 1, 2, \ldots, N$ individuals who are each observed at $t = 1, 2, \ldots, T$ time periods. Next, let $q_i$ be a measured time-stable variable, such as sex or race. We also let $x_{it}$ be a measured time-varying variable, such as delinquent peer exposure, whether an individual is married, or job attachment. Now, we let $\tau_i$ be a term that captures all of the time-invariant factors that generate criminal behavior that are not measured (i.e., persistent unobserved sources of heterogeneity). While $\tau_i$ is an error term (because it captures the effects of variables that are not measured), it is not a random-error term. The term $\varepsilon_{it}$ is a random-error term that captures all of the omitted variables that affect criminal behavior at different time periods. A key feature of this error term is that it is uncorrelated with $q_i$, $x_{it}$, and $\tau_i$. The true crime-generating process in this case can be written as,

$$y_{it} = \beta q_i + \delta x_{it} + \tau_i + \varepsilon_{it},$$

where $\beta$ and $\delta$ are parameters to be estimated.

Conventional fixed-effect estimators actually estimate the $\tau_i$ directly; the obvious cost of this procedure is that one is no longer able to estimate $\beta$. In the standard random-effects models commonly used in criminological research, each individual is assumed to draw a value of $\tau_i$ from a well-defined probability distribution. Random-effects estimators provide estimates of parameters associated with these probability distributions. For example, the GLS/LISREL framework allows researchers to investigate the covariances of the error terms for the outcomes $y_{it}$. From these covariances, one can recover the parameters of the distribution governing the random effects. The validity of this procedure, however, depends on a critical assumption: $\tau_i$ must be uncorrelated with both $q_i$ and $x_{it}$. If the stable unmeasured sources of heterogeneity (i.e., $\tau_i$) are correlated with $q_i$,
any conventionally obtained estimate of $\beta$ will be biased. Say, for example, that $q_i$ is an indicator of one's sex and also assume that poor early parenting practices (e.g., lack of caring, or lax, harsh, and erratic discipline) are criminogenic and that they are unmeasured. In this case, $\beta$ will not only indicate the effect of sex but it will also capture the effect of the part of sex that is correlated with poor parenting.

Similarly, if $\tau_i$ is correlated with $x_{it}$, the estimate of $\delta$ will also be biased. For example, if delinquent peer exposure is the time-varying predictor, $x_{it}$, and it is correlated with poor parenting, the estimate of $\delta$ will capture both the true effect of delinquent peer exposure and the part of poor parenting that is correlated with delinquent peer exposure. In both cases, the biases are the logical outcome of a flawed model specification (classic omitted variable bias).

We refer to this as the zero-covariance problem and our question is whether it is possible to do anything about it. Unfortunately, as we demonstrate below, there is a two-part answer to this question. First, it is impossible to avoid bias in $\beta$ when $q_i$ and $\tau_i$ are correlated with each other. Second, it is possible to reduce substantially the bias in $\delta$ when $x_i$ and $\tau_i$ are correlated with each other. As Nagin and Paternoster (1991; see also Horney et al., 1995) have pointed out, such correlations are quite plausible. Indeed, it is hard to envision realistic scenarios in which the zero-covariance problem could be ignored.

There are many theoretical reasons to expect a substantial positive association between unmeasured stable sources of criminality and measured covariates. A classic example, advanced by Hirschi (1969), is the contention that the frequently observed positive correlation between delinquent behavior and delinquent peers is spurious rather than causal. His position is that there is something about delinquents that leads them to seek out like-minded others—the process by which “birds of a feather flock together.”

Although we have illustrated this point with delinquent peers as a time-varying variable, the existence of nonzero covariance could create problems for inference about many other time-varying variables as well. Consider, for example, indicators of marital or job satisfaction. One can easily imagine a time-stable component of marital/job satisfaction that is also related to criminal offending, and of course, a time-stable component of offending that is related to marital/job satisfaction. Indeed, this is the very argument that Hirschi and Gottfredson (1995) make as to why reported “dynamic” effects are spurious. To assume that the covariance between unmeasured time-stable predictors of crime and measured time-varying predictors of crime is zero is unsatisfying precisely because this is the source of the theoretical controversy between static theorists and life-course dynamic theorists.
Further, as we show later, simply assuming that the covariance is zero when in fact it is not leads to estimates that overstate the impact of variables included in the model (both time-fixed and time-varying variables are affected). Notably, since time-varying variables can be biased, analyses that estimate important effects of time-varying covariates do not necessarily demonstrate that those effects are actually important. In our view, this is precisely the focus of Gottfredson and Hirschi’s critique of longitudinal research.

UNOBSERVED INITIAL CONDITIONS

Criminologists will often be interested in estimating a model that is slightly more complicated than Equation 1. A common alternative random-effects specification is to augment Equation 1 with a measure of prior criminal behavior, which we denote as \( y_{it-1} \). The model then takes the following form:

\[
y_t = \gamma y_{t-1} + \beta q_t + \delta x_t + \tau_i + \varepsilon_t,
\]

where \( \gamma \) is an estimated regression coefficient. All of the other parameters have the same meaning as in Equation 1. The estimated effect of prior offending estimated by models of the form in Equation 2 is of substantial interest. As Nagin and Paternoster (1991) suggest, such an estimate measures the collective impact of unobserved time-varying variables whose effects are not confined to only one period of observation. In other words, the effect of any unobserved time-varying variable whose effects endure for more than one period will be incorporated into the estimate of the effect of prior offending. Intuitively, if all such time-varying variables were observed and all sources of unobserved stable heterogeneity had been taken into account, the estimated effect of prior offending on current offending would vanish (so long as the model was properly specified).

Note that for valid estimation of Equation 2 using random-effects methods, the zero-covariance assumption is still required because \( q_t \) and \( x_t \) are still in the model. The zero-covariance assumption is not needed for lagged \( y \), however, because \( y_{it-1} \) is handled differently from the other time-varying explanatory variables. But correlation between \( y_{it-1} \) and \( \tau_i \) is still a problem, albeit in a different way.

The model requires that the offending process be observed at the beginning, the so-called initial-conditions assumption. In the first period when any individuals begin offending \( (t = 1) \), \( y_{it-1} \) will be equal to zero for all individuals. If this is true, \( y_{it-1} \) will be uncorrelated with \( \tau_i \), the unobserved measure of individual heterogeneity. As a result, there will be no confounding between lagged \( y \) and unobserved heterogeneity in this first period. The error term will include the full amount of unobserved heterogeneity.
Note that for subsequent periods, this will no longer be the case, because one expects that time-stable omitted variables, such as poor early parenting, will be correlated with subsequent offending once offending has started. Hsiao (1986:78–81) shows that the estimates of the effect of $y_{i,t-1}$ on contemporary $y$ will be consistent, provided the initial-conditions assumption has been met. If the initial-conditions assumption is not met, however, random-effects models will yield upwardly biased estimates of the effect of lagged $y$ and the estimates of the parameters associated with the probability of $\tau_i$ will be biased as well.

To provide a concrete illustration of the problem, assume we have collected four years worth of data that measure offending behavior at each year. Now, further assume that some members of the sample actually began to offend one year before the data collection effort began. In this case, the first observed time period is not the first time period in which individuals could have exhibited offending activity. Instead, it is the second such time period and there is no alternative but to base the analyses on the information in these four waves of data. The problem here, of course, is that the set of observed outcomes will include $y_{i3}, y_{i4}$, and $y_{i5}$, while the set of lagged outcomes will include $y_{i2}, y_{i3}$, and $y_{i4}$. Since the value of $y_{i,t-1}$ at the first observed time period is $y_{i1}$ and one cannot observe $y_{i1}$, one is forced to treat $y_{i3}$ as the first observed time period. The lagged outcome at the first observed time period is, therefore, $y_{i2}$. Unfortunately, $y_{i2}$ will vary from individual to individual. In most cases, unless one is willing to assume that there is no correlation between $y_{i2}$ and $\tau_i$, the failure to observe initial conditions guarantees that parameter estimates from the random-effects model will be biased and inconsistent.

Consider Nagin and Paternoster’s (1991) “state-dependence” argument and Sampson and Laub’s (1992, 1993, 1997) life-course theory. Both of these frameworks hypothesize that prior offending experience has causal implications for future offending activity. The idea of state dependence and the notion of the “cumulative continuity of disadvantage” predict this kind of causal sequence. The problem that initial conditions pose for discerning cause and effect within the context of continuity and change in criminal offending was first raised by Nagin and Paternoster (1991:184):

In most panel data sets used by criminologists, including the panel used in this analysis, an appreciable portion of the sampled population has already engaged in delinquency prior to the initial year of the panel. Because, by definition, the panel does not include data on the individual’s circumstances prior to its initial period, the state-dependence versus heterogeneity interpretation of pre-initial involvement is not easily sorted out.
have often not directly confronted the initial-conditions problem and have either ignored it or assumed that the initial conditions were exogenous. We examine below the extent to which this is a critical problem that cannot be assumed away when the model includes lagged measures of the dependent variable. We also implement a partial but useful solution to this problem.

In sum, it may appear that Gottfredson and Hirschi's expressed skepticism about the capability of multivariate analysis of longitudinal data in distinguishing cause from effect is well-grounded. Random-effects panel analysis is fraught with potential pitfalls, and even with complex statistical treatment of the data it may be impossible to distinguish correlation from causality. Unless one is observing the onset of the causal process, or can make strong assumptions about its exogeneity, estimates of key parameters will be biased.

Recall, however, that even if the initial-conditions problem is resolved, not all of the bugs are out of the basement. We have previously discussed the problem that arises when there is a stable component of an explanatory variable that is correlated with the error term in the outcome variable. This problem will arise whenever stable individual crime-producing characteristics are correlated with variables such as delinquent peers, marital satisfaction, or job stability—in other words—many of the life-course events in which dynamic theorists are most interested. Unfortunately, we have thus far neither shown the magnitude of the problem (i.e., the extent to which parameter estimates are biased) nor suggested any reasonable solutions to these problems. In the next section we illustrate how severe the problems of initial conditions and zero covariance are for accurately estimating the structural parameters of random-effects panel models. Our platform for doing this is a series of simulations in which readers can see quite clearly the difference between the true and estimated parameter values when assumptions are violated. After documenting the magnitude of the bias when the assumptions of random-effects panel models are violated, we turn to a discussion of what we think are useful ways to address the initial-conditions and zero-covariance problems. We illustrate their application through additional simulations and a reanalysis of the data from a paper two of the current authors recently published, in which these problems were squarely at issue.

ASSESSING THE MAGNITUDE OF THE PROBLEM

The major points we wish to make can be illustrated through the use of a relatively simple set of simulations. In this section, our objective is to document the direction and magnitude of the estimation bias one can expect to encounter when the initial-conditions and zero-covariance assumptions are violated, under conditions similar to what is usually
encountered in the field. To achieve this objective, we simulated data for analysis with three kinds of statistical models commonly estimated by criminologists: (1) covariance structure (LISREL) (or, generalized least squares) models for continuous normally distributed outcomes; (2) probit models for binary outcomes; and (3) negative binomial models for event-count outcomes. Our simulations in each case involved samples of 500 individuals followed for five periods. The sample size and number of periods were chosen to approximate the size and nature of some of the more well-known panel data sets, like the National Youth Survey, the Rochester Youth Development Study, the Pittsburgh Youth Study, the Denver Youth Study, Farrington’s Cambridge data, and the Gluecks’ data set recently analyzed by Sampson and Laub (1993).

The data were generated with five key components generally believed to exist in criminological data: (1) an observable, time-stable, individual-specific characteristic, like race or sex, written as $q_i$; (2) an unobservable, time-stable, individual-specific characteristic, like intelligence or temperament, written as $\tau_i$; (3) an observable, time-varying (dynamic) component, like marriage, employment, or delinquent peer exposure, written as $x_{it}$; (4) a time trend that allows for shifts in the overall rate of offending that correspond to the passage of time; and (5) a measure of offending that is causally dependent on observed and unobserved heterogeneity, the time trend, $t$, a time-varying, individual-specific component, and offending in the previous period, $y_{it-1}$. The $q_i$ and $\tau_i$ were drawn from the standard unit bivariate normal distribution and allowed to be correlated at $\rho = 0.3$. The time-varying covariate, $x_{it}$, was generated according to two protocols. First, we generated $x_{it}$ by a first-order autoregressive process that depends on observed heterogeneity only. Then, we generated $x_{it}$ by a first-order autoregressive process that depends on observed and unobserved heterogeneity. The correlation between individual heterogeneity and the dynamic component of the model is a fundamental issue in the current debate between dynamic and static models of criminal offending. For simplicity, the coefficients on all variables except the time-trend indicator, $t$, were set to 0.5. The time-trend indicator coefficient was set to 0.1.

The three outcomes were generated by using the so-called “acceptance-rejection method” (see e.g., Fishman 1996:171–172). Briefly, implementation of this method requires access to a uniform random number generator and a program in which probability density or mass functions can be easily written down. Implementation details can be found in Algorithm AR, described by Fishman (1996:172). For the LISREL/GLS simulations, each simulated realization of the dependent variable is a draw from the normal

\footnote{This ensures that there will be significant correlation between residuals of different time periods for the same individual.}
distribution. For the probit simulations, each simulated realization of the dependent variable is a draw from the Bernoulli distribution. For the event-count simulations, each simulated realization of the dependent variable is a draw from a beta mixture of negative binomial distributions.

The top panel of Table 1 presents summary results for 100 simulations in which both the initial-conditions and zero-covariance assumptions are met. Because the observed time-stable covariate, $q_i$, is positively correlated with the unobserved heterogeneity, $\tau_i$, the average estimate of its effect is positively biased due to omitted variable bias. The estimates of the effect of the lagged outcome variable and the time-varying covariate, $x_{it}$, are very close to the values of 0.5, which were used to generate the data. These results suggest that when one estimates random-effect probit models under ideal conditions, most parameters (including the effect of $y_{it}$) can be consistently estimated. This result differs from the conclusion of Heckman (1981), who reported bias in the random-effect probit estimator under ideal conditions. In the probit model, the estimate of the variance of the unmeasured population heterogeneity, $\rho$, is also on target.

The middle panel of Table 1 presents the summary results of simulations in which the zero-covariance assumption is still correct, but the first of the five discrete periods is dropped from the data for each of the 500 observations. This induces a violation of the initial-conditions assumption. In all three sets of simulations, the estimated effect of offending in the previous period is positively biased. Moreover, the amount of bias evident in these simulations is not trivial. In the LISREL/GLS simulations, the estimated effect of the lagged outcome variable was 74% larger than the value of 0.5 that was used to generate the data. We obtained large biases for the estimated effect of the lagged outcome in the probit (142%) and negative binomial models (45%) as well. This result is consistent with Heckman's (1981) argument that violation of the initial-conditions assumption will lead to upwardly biased inferences about the magnitude of the effect of the lagged outcome variable. According to Heckman, this bias can be directly attributed to the confounding of unobserved heterogeneity and prior offending. Unfortunately, when one fails to observe the initial conditions, standard random-effects models are incapable of resolving this problem.

Because violation of the initial-conditions assumption generates substantial bias in the estimated effects of the lagged outcome variable, an
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Initial Conditions</th>
<th>Initial Conditions</th>
<th>Initial Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Observed</td>
<td>Not Observed</td>
<td>Not Observed and cov(x,t) &gt; 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LISREL</td>
<td>Probit</td>
<td>Negative Binomial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>σ</td>
<td>E</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.5</td>
<td>-1.501 .126</td>
<td>2.669 .941</td>
<td>-1.317 .169</td>
</tr>
<tr>
<td>Time Index</td>
<td>0.1</td>
<td>.100 .030</td>
<td>.101 .016</td>
<td>.060 .041</td>
</tr>
<tr>
<td>Time 1 Dummy</td>
<td>-1.4</td>
<td>-1.403 .066</td>
<td>-1.456 .070</td>
<td>1.832 .065</td>
</tr>
<tr>
<td>Time 2 Dummy</td>
<td>-1.3</td>
<td>-1.283 .067</td>
<td>-1.456 .070</td>
<td>1.832 .065</td>
</tr>
<tr>
<td>Time 3 Dummy</td>
<td>-1.2</td>
<td>-1.188 .069</td>
<td>-1.456 .070</td>
<td>1.832 .065</td>
</tr>
<tr>
<td>Time 4 Dummy</td>
<td>-1.1</td>
<td>-1.085 .069</td>
<td>-1.456 .070</td>
<td>1.832 .065</td>
</tr>
<tr>
<td>Time 5 Dummy</td>
<td>-1.0</td>
<td>-0.997 .082</td>
<td>-1.456 .070</td>
<td>1.832 .065</td>
</tr>
<tr>
<td>Time-Stable Covariate (q)</td>
<td>0.5</td>
<td>.803 .058</td>
<td>.790 .092</td>
<td>.795 .066</td>
</tr>
<tr>
<td>Time-Varying Covariate (x)</td>
<td>0.5</td>
<td>.499 .025</td>
<td>.498 .043</td>
<td>.503 .024</td>
</tr>
<tr>
<td>Lagged Outcome</td>
<td>0.5</td>
<td>.502 .020</td>
<td>.505 .119</td>
<td>.493 .058</td>
</tr>
<tr>
<td>ρ</td>
<td>0.5</td>
<td>.473 .052</td>
<td>.473 .052</td>
<td>.473 .052</td>
</tr>
<tr>
<td>α</td>
<td>—</td>
<td>—</td>
<td>75.898 92.722</td>
<td>31.075 40.293</td>
</tr>
<tr>
<td>β</td>
<td>—</td>
<td>—</td>
<td>1.117 .114</td>
<td>1.354 .213</td>
</tr>
<tr>
<td>Endogenous Covariance</td>
<td>0.9</td>
<td>.869 .092</td>
<td>1.014 .045</td>
<td>1.014 .045</td>
</tr>
<tr>
<td>ψ_{11}</td>
<td>1.8</td>
<td>1.838 .120</td>
<td>1.367 .084</td>
<td>1.367 .084</td>
</tr>
<tr>
<td>ψ_{22}</td>
<td>1.8</td>
<td>1.820 .098</td>
<td>1.192 .084</td>
<td>1.192 .084</td>
</tr>
<tr>
<td>ψ_{33}</td>
<td>1.8</td>
<td>1.833 .117</td>
<td>1.070 .066</td>
<td>1.126 .076</td>
</tr>
<tr>
<td>ψ_{44}</td>
<td>1.8</td>
<td>1.839 .124</td>
<td>1.138 .078</td>
<td>1.138 .078</td>
</tr>
<tr>
<td>ψ_{55}</td>
<td>1.8</td>
<td>1.836 .114</td>
<td>1.138 .078</td>
<td>1.138 .078</td>
</tr>
</tbody>
</table>

**NOTE:** ρ indexes the correlation between the error terms of the various waves for the probit model. The α and β terms govern the beta distribution from which the negative binomial random effects are drawn. The Ψ symbols represent the variances of the dependent variable at each wave of the panel, while the endogenous covariance captures the covariances between the error terms of the various waves. These terms are associated with the LISREL/GLS specification. We use the term E to denote the expected value of the parameter (averaged across the set of simulation results), and the term σ is used to denote the standard deviation.
interesting question is whether this bias also contaminates other parameter estimates. In general, the answer to this question appears to be yes. In the GLS/LISREL and probit frameworks, our estimates of the effects of $q_i$ and $x_{it}$ were negatively biased. In the negative binomial framework, the estimated effect of $q_i$ is positively biased while the estimated effect of $x_{it}$ is essentially on target. In the probit model, there also is evidence of negative bias in the estimated effect of unobserved population heterogeneity since the lagged outcome absorbs much of the variability attributed to static factors. The estimated coefficient of 0.060 for $\rho$ underestimates the true value of .50 by 88%.

The far right panel of Table 1 shows the consequences of violating both the initial-conditions and zero-covariance assumptions. In each of our three models, parameter estimates associated with the time-stable covariate $q_i$ and the lagged outcome variable continue to be biased. Moreover, the estimates associated with $x_{it}$ are now positively biased because $x_{it}$ and $\tau_i$ are positively correlated. Finally, while the true value of unobserved heterogeneity in the probit model is .50, the mean estimated value of $\rho$ is only .002 when both assumptions of the random-effects model are violated.

The bias also propagates to other parameter estimates when both assumptions are violated. In the GLS/LISREL framework, there is a 63% bias in the estimate of the lagged outcome. In the probit and event-count frameworks, the bias is 138% and 97%, respectively. The positive bias for the time-varying dynamic covariate $x_{it}$ is 8% in the GLS/LISREL model and approximately 13% in the probit and event-count frameworks. The negative bias for the variance component, $\rho$, in the probit model is 99.6%. The reader should notice that the positive bias in both $x_{it}$ and $y_{it-1}$ and the negative bias for the estimated importance of unmeasured time-stable crime-producing factors, $\rho$, in the probit model is in the direction that exaggerates the importance of dynamic factors and understates the importance of static factors.

The importance of this last point cannot be overemphasized. If this had been an actual analysis with unobserved initial conditions, the model specification in the last column of Table 1 would lead us to numerous incorrect conclusions. First, we would have concluded that there is no unobserved heterogeneity in the data, even though this is clearly not the case. The dynamic factor captures all of the variation that should be attributed to static factors. Second, we would have grossly overestimated the importance of the effect of the lagged outcome. Third, the biases associated with $q_i$ and $x_{it}$, while smaller, are also substantively important. Taken as a whole, the estimates from these models bear little resemblance to the values that were actually used to generate the data. Next, we address the
question of whether it is possible to improve on the poor performance of the estimators in Table 1.

PROPOSED SOLUTIONS

INITIAL CONDITIONS

Frequently, it simply is not possible to observe the initial state of some causal process. This results in a violation of the initial-conditions assumption. When one is working with continuous, normally distributed outcomes, there are many options available to deal with the initial-conditions problem. Among the possibilities are conventional fixed-effect estimators and alternative random-effects estimators that make different assumptions about the relationship between the unobserved heterogeneity, \( \tau_i \) and \( \psi_{io} \). These options are discussed in detail in Kessler and Greenberg (1981: Ch. 2, 3, and 7) and Hsiao (1986:25–96). Our experience with these estimators suggests that they perform quite well in situations in which one can assume that the outcome variable is normally distributed. Moreover, these estimators are simple to implement with well-known software packages, such as SAS, Gauss, or S+. Unfortunately, as discussed above, most outcome variables studied by micro-level criminological researchers are either binary or event counts, and methods that take the discrete features of the data into account are necessary (Hsiao 1986:154–180). We focus the balance of our attention on remedies that can be used with qualitative outcomes.

When initial conditions are unobserved and cannot be assumed to be exogenous, Heckman (1981) provides the likelihood function that can be used to estimate the probability of offending in the first observed period. With current tools, this likelihood function is virtually intractable. Fortunately, he also provides a much simpler approximation from the available data. In the case at hand, the goal of the approximation is to estimate the probability of offending in the first period net of the individual’s underlying criminal propensity. In other words, if one cannot actually observe the initial conditions and cannot assume that it is exogenous, one can estimate the approximate state at the beginning of the series—that is, one can estimate the initial conditions from available data (Heckman 1981:188; Hsiao 1986:169–172). Since this solution approximates the state of initial conditions, we call it the Heckman approximation.

Implementation of the Heckman approximation is fairly straightforward. Exogenous variables at the first wave are used to estimate the reduced-form marginal probability of offending at that wave. This probability is estimated from an auxiliary probit regression model, and the estimated probability of offending at the first wave of data from this model is used to “stand in” for the actual observed value of criminal activity.
Essentially, then, one takes the predictor variables from the first available wave, identifies exogenous variables to predict criminal offending, estimates a probit regression model for criminal offending for that wave, and then uses this estimated probability in subsequent panel analyses of the data in place of the observed value for the first wave.

In our example from above in which we did not observe the causal process until the second time period, assume that we were able to locate an exogenous variable $q_i$ that predicts participation in criminal activity. In the Heckman approximation, we would then take this exogenous variable and estimate the following probit equation for the probability of offending at the second wave:

$$p(y_{i2} = 1) = \Phi[\eta + \theta q_i],$$

where $\eta$ and $\theta$ are estimated probit coefficients and $\Phi[.]$ is the standard normal cumulative distribution function. Next, we enter this estimated probability of time 2 offending into the equation for time 3 offending. Thus, the probit equation for time 3 offending takes the following form:

$$y_{i3}^* = \alpha_3 + \beta q_i + \delta x_i + \gamma Pr(y_{i2} = 1) + \tau_i + \epsilon_{i3},$$

where $p(y_{i2} = 1)$ is obtained from Equation 3, $\alpha_3$ is the value of $y_{i3}^*$ when all of the covariates are equal to zero, and

$$y_{i3} = \begin{cases} 0 & \text{if } y_{i3}^* \leq 0 \\ 1 & \text{if } y_{i3}^* > 0 \end{cases}.$$

To implement the Heckman approximation, it is appropriate to use variables that are available in the first observed time period. In practice, an auxiliary probit model can never include all of the characteristics that generate variation in offending propensity and, therefore, not all of the bias will be eliminated by this procedure. Still, the bias will be smaller with the Heckman approximation than if the initial-conditions problem is simply ignored.10

10. One reviewer suggested that the initial-conditions problem can be ameliorated by obtaining data sets in which the number of time periods is large. This was an intriguing suggestion to us. In order to investigate it, we conducted a Monte Carlo simulation experiment in which we generated 17 periods of data in a form that we could estimate the random-effect probit model. In this simulation, the true effect of the lagged outcome was 0.3. We then dropped the first period entirely, and the second period outcome was used only to supply the lagged outcome for the third period. We then estimated the random-effect probit model using periods 3–7. The estimated effect of the lagged outcome was 0.8, evidencing substantial bias. We then estimated the probit model using periods 3–12. In this model, the estimated effect of the lagged outcome was 0.52, evidencing nonignorable but less substantial bias. Next, we estimated the probit model using periods 3–17. The estimated effect of the lagged outcome was 0.32, virtually on target. In the model in which the initial conditions were observed, the
ZERO COVARIANCE

Recall that the zero-covariance problem is essentially a specification error that is caused by omitted variable bias. That is, some of the variation in the predictor variables could be covarying with stable, unmeasured differences in the tendency to offend. To the extent that such covariation exists, one can expect parameter estimates to be biased and inconsistent. Unfortunately, for time-invariant covariates, such as sex and race, very little can be done to resolve this problem. For time-varying covariates, such as delinquent peers, job stability, or marital satisfaction, however, it may be possible to gain some improved accuracy in these situations.

A general solution to this problem involves the decomposition of $x_{it}$ into two distinct components. The first component is the time-stable part of $x_{it}$, and the second component is the part of $x_{it}$ that changes its values over time. The most promising approach available for accomplishing the objective of separating the between-individual variation in $x_{it}$ from the within-individual variation in $x_{it}$ in a random-effects framework is the procedure suggested by Bryk and Raudenbush (1992), employed in a recent paper by Homey et al. (1995). We continue to use delinquent peer exposure as our example of a time-varying independent variable.

The first, time-stable, component of the delinquent peer variable is each individual's level of delinquent peer exposure averaged across all of the waves of data. This captures between-individual variation in delinquent peer exposure. The second, time-varying, component of the delinquent peer variable is calculated separately for each time period within each individual person. For each person, we take the difference between the overall average delinquent peer exposure and the delinquent peer exposure at time $t$.

In order to separate the between-individual component of a time-varying covariate like delinquent peers from the within-individual component, the following steps should be followed: (1) for each individual, calculate the average level of delinquent peer exposure across all waves:

estimated effect of the lagged outcome was 0.28. On the basis of this evidence, we conclude that this reviewer's intuition that the initial-conditions problem could be adequately handled with more waves of data was correct. For researchers fortunate to have access to data sets with numerous time periods (bias was ignorable with about 15 time periods), then, there is another way to tackle the initial-conditions problem. However, we must note that few publicly available data sets in criminology will allow researchers to appeal to this "big T" solution. Most panel data sets in criminology have a far more limited number of observation periods. In that instance, the solution we propose here is going to be the solution of choice. Although alternative data collection strategies promise "big T" data sets (like diary methods), these are still infrequent in our field.
(2) at each wave, calculate the deviation between the period-specific delinquent peer exposure and the level of delinquent peer exposure for that individual averaged across all periods obtained from step (1):
\[
\Delta x_{it} = x_{it} - \bar{x}_i
\]
and include the deviation score, \(\Delta x_{it}\), in the model; and (3) add the overall average level of delinquent peer exposure, \(\bar{x}_i\), to the specification in Equation 2. In essence, the time-varying predictor variable is now "individual" mean-centered and the correlation between \(\Delta x_{it}\) and the error term for \(y_{it}\) is zero.

11. This last step is taken because of the properties of linear fixed-effects estimators. To see this, consider the following fixed-effect equation in a linear model:
\[
y_{it} = \alpha + \beta x_{it} + \tau_i + \varepsilon_{it}.
\]
If we take the within-individual mean across all time periods for each individual, we are left with the following equation:
\[
\bar{y}_i = \frac{T}{T} \alpha + \beta \bar{x}_i + \tau_i + \bar{\varepsilon}_i,
\]
which reduces to
\[
\bar{y}_i = \alpha + \beta \bar{x}_i + \tau_i
\]
because the average error is 0 by definition. If we subtract Equation b from Equation a, we are left with Equation c, which no longer includes the fixed effect, but still allows us to estimate beta, the coefficient of interest:
\[
y_{it} - \bar{y}_i = \beta (x_{it} - \bar{x}_i) + \varepsilon_{it}.
\]
This approach is often used in the continuous framework to avoid estimating a dummy variable for each individual.

Now, as discussed above, the fixed-effect framework is not always satisfying when the dependent variable is a qualitative outcome. The result is reliance on the random-effects model. But what Bryck and Raudenbush have done is start from the fixed-effect framework in the linear model, in essence Equation c. From this perspective, the problem is \(\bar{y}\), which is not particularly meaningful in the discrete case because the value of \(y\) is arbitrarily set to 1 and 0 to represent qualitatively different states.
Equation b, though, gives us a substitute for \(\bar{y}\) that is entirely in terms of an intercept, \(\bar{x}\) (which is meaningful assuming \(x\) is continuous) and \(\tau\). So we substitute Equation b into Equation c and solve for \(y_{it}\). Simplifying gives us Equation d:
\[
y_{it} = \alpha + \beta (x_{it} - \bar{x}_i) + \beta \bar{x}_i + \tau_i + \varepsilon_{it}.
\]
Notice the similarity between this equation and Equation 1 in the text. Provided the zero-covariance assumption is met, this can be estimated using the random-effect estimator. The zero-covariance assumption is not met for \(\bar{x}\), but it is met for the within mean estimate for \(x\). Despite the bias on the coefficient for \(\bar{x}\), we should get an unbiased estimate for the within-mean \(x\), since there is no correlation between the within-mean \(x\) and \(\bar{x}\).
Put another way, this procedure provides an estimate of the effect of $\Delta x_{it}$ on $y_{it}$ by separating the within- and between-individual variation in $x_{it}$ while including them separately in the model. The between-individual relationship will be captured by the individual's mean averaged over all time periods. Because it does not vary over waves for the individual, it is necessarily uncorrelated with any within-individual variation over time on $y_{it}$ and can only be related to the mean level of $y_{it}$ for an individual over time. Correspondingly, the deviation between the individual's score at time $t$ and the average score over all time periods is necessarily uncorrelated with the individual mean on $y_{it}$ because each individual's mean deviation score is 0. This deviation score is, then, devoid of any correlation with $x_{it}$ and any unmeasured stable individual difference.12

For one time-varying independent variable and a lagged dependent variable, the probit specification would be:

$$y_{it} = \alpha_i + \beta x_{it} + \gamma y_{it-1} + \delta_{wit} \delta_{wit} + \delta_{wit} \Delta x_{it} + \tau_i + \epsilon_i, \tag{5}$$

where $\delta_{wit}$ is the within-individual effect of the time-varying independent variable (again, say, delinquent peer exposure) and $\delta_{wit}$ is the between-individual effect of the time-varying independent variable.

The major advantage of this mean-centering procedure is that it divorces the time-stable portion of delinquent peer exposure from that part of delinquent peer exposure that varies over the several survey waves (Liker et al., 1985). It thereby guarantees that the deviation scores will be uncorrelated with unmeasured and time-stable sources of criminal propensity. We are unsure about the full range of merits and demerits associated with step (3) of this procedure. We have conducted all of our subsequent analyses both with and without this step, but its appropriateness should be investigated further in future research. We would note, however, that if it is included in the model, the effect of the overall mean for the time-varying independent variable, $\delta_{biv}$, (from Equation 5) should not be interpreted.

Now, the question remains, do the proposed solutions to the initial-conditions and zero-covariance problems (i.e., the Heckman approximation and the mean-centering of the time-varying independent variable) really reduce the bias in the estimated parameters of a random-effects panel model? To answer this question we appeal first to simulation results and then we reexamine some results previously reported in this journal.

The problem with this approach is that it is based on the linear model, and we are discussing the random-effect estimate within the nonlinear framework, in which Equation b no longer holds. So the theoretical basis for this process also no longer holds. Yet, as we will show below in our simulations, this "fix" provides biased but good estimates of the desired coefficients.

12. We acknowledge the assistance of Wayne Osgood in helping us understand this issue.
SIMULATION RESULTS

In this section, we use simulations to investigate the performance of the suggested remedies to the estimation problems that arise when the initial-conditions and zero-covariance assumptions are violated. Our reported results are restricted to the probit model. We note, however, that we obtained similar results with the negative binomial estimator for event-count outcomes.

Table 2 presents three sets of probit simulation results. In the first panel of Table 2, we apply the Heckman approximation to the case in which initial conditions are not observed and the correlation between the time-varying independent variable, $x_{it}$, and stable unobserved heterogeneity $\tau_i$ is zero. The average parameter estimates for this simulation are very close to those for the ideal case (panel 1 of Table 1). The single exception to this general pattern is the negatively biased estimate of the effect of $y_{it-1}$. Still, the bias in this estimate is dramatically smaller (14%) than the bias that results from simply ignoring the initial conditions problem (142%).

In addition, inferences based on this estimate will tend to be conservative since they understate the dynamic effect of the lagged outcome variable. It is also worth noting that the standard error of the effect of the lagged outcome variable is twice as large as what is observed with the naive estimator. In sum, while the Heckman approximation yields an improved average estimate of the effect of the lagged outcome, this improvement comes at the cost of increased sampling variability of that estimate.

In the second panel of Table 2, we turn to the case in which initial conditions are violated and there is a positive correlation between the time-varying covariate and unobserved time-stable individual heterogeneity. The probit simulations in this panel incorporate the Heckman approximation for initial conditions and the division of the time-varying covariate into between- and within-subject components. As expected, the introduction of correlation between the time-stable unobserved heterogeneity and the time-varying covariate leads to some deterioration in the accuracy of the parameter estimates. The simulation results indicate that the variance component that measures the importance of unmeasured time-stable crime-generating processes, $\rho$, the effect of the lagged outcome, $y_{it-1}$, and the effect of the time-invariant predictor variable, $q_i$, are all underestimated, while the effect of the between-individual component of the time-varying covariate is dramatically overestimated. Apparently, the between-individual component of the time-varying covariate captures a great deal of the static information in the data. The average estimate of the within-individual component, however, is very close to the true parameter value. As Horney et al. (1995) have indicated, if one's substantive interest is in...
Table 2. Probit Simulations with Corrections for Initial Conditions and Zero-Covariance Problems

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>( \hat{E} )</th>
<th>( \sigma )</th>
<th>( \hat{E} )</th>
<th>( \sigma )</th>
<th>( \hat{E} )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.5</td>
<td>-1.518</td>
<td>.250</td>
<td>-1.542</td>
<td>.030</td>
<td>-1.486</td>
<td>.274</td>
</tr>
<tr>
<td>Time Index</td>
<td>0.1</td>
<td>.101</td>
<td>.057</td>
<td>.105</td>
<td>.067</td>
<td>.093</td>
<td>.064</td>
</tr>
<tr>
<td>Time-Stable Covariate (q)</td>
<td>0.5</td>
<td>.840</td>
<td>.154</td>
<td>.345</td>
<td>.119</td>
<td>1.126</td>
<td>.195</td>
</tr>
<tr>
<td>Time-Varying Covariate (x)</td>
<td>0.5</td>
<td>.517</td>
<td>.072</td>
<td>1.467</td>
<td>.225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Level of ( x )</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period-Specific Deviation of ( x )</td>
<td>0.5</td>
<td>.428</td>
<td>.255</td>
<td>.347</td>
<td>.257</td>
<td>.464</td>
<td>.238</td>
</tr>
<tr>
<td>Lagged Outcome</td>
<td>0.5</td>
<td>.516</td>
<td>.098</td>
<td>.367</td>
<td>.115</td>
<td>.644</td>
<td>.079</td>
</tr>
</tbody>
</table>

NOTE: \( \rho \) indexes the correlation between the error terms of the various waves for the probit model. We use the term \( \hat{E} \) to denote the expected value of the parameter (averaged across the set of simulation results), and the term \( \sigma \) is used to denote the standard deviation.
the effects of time-varying covariates, this method appears to perform quite well.

Sometimes, of course, investigators wish to interpret the effect of the lagged outcome variable \( y_{i,t-1} \) as well as the effects of time-varying covariates. The method in the second panel of Table 2 does not perform well in this context. The third panel of Table 2 presents results of the simulation in which the between-individual component of the time-varying covariate was omitted from the model. These simulation results show that this omission leads to an improved estimate of the effect of the lagged outcome at the cost of a small increase in the bias of the estimated within-individual effect of the time-varying covariate. The results also indicate that the effect of the time-stable covariate, \( q_i \), is once again positively biased.

In sum, these simulations point to the conclusion that the covariance problem can be largely overcome. However, there is no complete remedy for the initial-conditions problem especially when the covariance problem also exists. Nevertheless, Heckman’s approximation produces much more accurate estimates of the effect of the lagged outcome variable provided appropriate predictor variables for estimating the probability of offending at the first wave are available.

REANALYSIS OF PREVIOUS WORK

The problem of initial conditions and zero covariance arose in a panel analysis reported by two of the current authors in a recent article in this journal (Paternoster and Brame, 1997). There were two substantive issues in that article: (1) Drawing on a theoretical debate between Hirschi and Gottfredson (1995) and Sampson and Laub (1995; see also Sampson and Laub 1993:197–200), we asked whether time-varying independent variables indexing prior offending activity and delinquent peer exposure were associated with participation in and frequency of criminal activity net of controls for observed and unobserved heterogeneity. (2) Drawing on the debate between general and typological theorists (Gottfredson and Hirschi, 1990; Moffitt, 1993; Patterson and Yoerger, 1993), we asked whether the effects of prior offending activity and delinquent peer exposure differed between high and low antisocial propensity groups.

The data to examine these issues were obtained from the first five waves of the National Youth Survey (NYS). We selected only the two youngest

---

13. Because of space limitations, we present only a brief overview of our earlier study. Interested readers can consult the original article for greater detail.

14. The original study included both a behavior- and an attitude-based measure of antisocial propensity, and the categorization into high or low antisocial propensity groups was based on the use of different cutoff points.
age cohorts, those youth who were 11 or 12 years old at the first wave (N = 509). After deleting cases that did not have sufficient data for the procedures used in the analysis, the sample comprised 471 individuals. A subsample of 390 cases had complete data at all four remaining waves of data collection. We conducted all of the analyses reported in that article with the entire sample of 471 cases and with the subsample of 390 cases that had complete information.

In favor of the position of dynamic theorists, the original findings suggested that the effects of delinquent peer exposure and prior offending activity exerted strong effects on criminal activity after controls for observed and unobserved individual differences had been taken into account. With respect to the debate on general and typological theories, we concluded that there was no evidence of important between-group variation in these effects.

In sum, our original analysis found strong support for general theories that anticipate important effects for time-varying variables. The parameter estimates in our original analysis were obtained from random-effects estimators (probit and negative binomial). For purposes of comparison, Table 3 restates the results of our original Table 4, and Table 4 restates the results of our earlier Table 6. In the following section, we revisit our previous results after employing the Heckman approximation for initial conditions and mean-centering the time-varying independent variable (delinquent peer exposure). In so doing, we attempt to provide better solutions to the problem of initial conditions and zero covariance than we employed in the original study.

As discussed earlier, when initial conditions are not observed, Heckman's approximation can be used to estimate the probability of offending at the first wave. This probability is estimated from an auxiliary probit regression model. We identified six variables in the data that we thought were clearly exogenous: (1) respondent's sex; (2) respondent's race; (3) whether the family receives public assistance; (4) whether the father was unemployed; (5) the age of the respondent's youngest parent; and (6) the educational level of the respondent's most well-educated parent. The specification of this equation was

\[ p(y_{it} = 1) = \Phi[\eta + \theta q_i], \]

15. Even after selecting on these two youngest age groups, we cannot plausibly claim that we have observed the initial conditions. Approximately 18% of these youth committed at least one offense during the first wave of data collection.

16. The analyses reported in our 1997 study were based on the sample of 471 cases. When cases with any missing data were excluded, however, all of the analyses in that study produced virtually identical results. In the current reanalysis, some of the results depend more heavily on cases with missing data.
Table 3. Baseline Probit and Negative Binomial Models Under Original Specification

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>z Ratio</th>
<th>Estimate</th>
<th>z Ratio</th>
<th>Estimate</th>
<th>z Ratio</th>
<th>Estimate</th>
<th>z Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Probit Participation Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude Propensity Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex = Male</td>
<td>.389</td>
<td>3.86</td>
<td>.379</td>
<td>3.86</td>
<td>.458</td>
<td>3.24</td>
<td>.461</td>
<td>3.21</td>
</tr>
<tr>
<td>Race = Nonwhite</td>
<td>.016</td>
<td>0.15</td>
<td>.023</td>
<td>0.21</td>
<td>.230</td>
<td>1.54</td>
<td>.229</td>
<td>1.44</td>
</tr>
<tr>
<td>Time Index</td>
<td>-.118</td>
<td>2.46</td>
<td>-.118</td>
<td>2.47</td>
<td>-.120</td>
<td>1.86</td>
<td>-.112</td>
<td>1.71</td>
</tr>
<tr>
<td>Any Missing Waves</td>
<td>.238</td>
<td>1.66</td>
<td>.236</td>
<td>1.65</td>
<td>.030</td>
<td>0.20</td>
<td>.042</td>
<td>0.27</td>
</tr>
<tr>
<td>Prior-Wave Delinquency</td>
<td>.837</td>
<td>6.84</td>
<td>.862</td>
<td>7.11</td>
<td>.959</td>
<td>6.03</td>
<td>.984</td>
<td>6.28</td>
</tr>
<tr>
<td>High-Propensity Indicator</td>
<td>.406</td>
<td>2.69</td>
<td>.332</td>
<td>2.22</td>
<td>.445</td>
<td>2.59</td>
<td>.505</td>
<td>3.12</td>
</tr>
<tr>
<td>( \rho (\sigma^2_i/(\sigma^2_i+\sigma^2_v)) )</td>
<td>.080</td>
<td>0.61</td>
<td>.068</td>
<td>0.51</td>
<td>1.829</td>
<td>3.95</td>
<td>1.928</td>
<td>3.65</td>
</tr>
<tr>
<td>( \alpha )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.538</td>
<td>2.19</td>
<td>3.977</td>
<td>2.08</td>
</tr>
<tr>
<td>( \beta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-531.99</td>
<td>-533.30</td>
<td>-1093.88</td>
<td>-1093.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Negative Binomial Frequency Models**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>z Ratio</th>
<th>Estimate</th>
<th>z Ratio</th>
<th>Estimate</th>
<th>z Ratio</th>
<th>Estimate</th>
<th>z Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude Propensity Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior Propensity Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** \( \rho \) indexes the correlation between the error terms of the various waves for the probit model. The \( \alpha \) and \( \beta \) terms govern the beta distribution from which the negative binomial random effects are drawn.
Table 4. Probit and Negative Binomial Models with Interaction Terms Under Original Specification

<table>
<thead>
<tr>
<th>Variable</th>
<th>Probit Participation Models</th>
<th>Negative Binomial Frequency Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attitude Propensity Measure</td>
<td>Behavior Propensity Measure</td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>z Ratio</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.875</td>
<td>14.11</td>
</tr>
<tr>
<td>Sex = Male</td>
<td>0.389</td>
<td>3.85</td>
</tr>
<tr>
<td>Race = Nonwhite</td>
<td>0.017</td>
<td>0.16</td>
</tr>
<tr>
<td>Time Index</td>
<td>-1.118</td>
<td>2.48</td>
</tr>
<tr>
<td>Any Missing Waves</td>
<td>0.232</td>
<td>1.61</td>
</tr>
<tr>
<td>Delinquent Peer Exposure</td>
<td>0.922</td>
<td>9.17</td>
</tr>
<tr>
<td>Prior-Wave Delinquency</td>
<td>0.878</td>
<td>6.67</td>
</tr>
<tr>
<td>High-Propensity Indicator</td>
<td>0.355</td>
<td>0.85</td>
</tr>
<tr>
<td>Interaction Terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propensity x Del. Peer Exp.</td>
<td>0.069</td>
<td>0.33</td>
</tr>
<tr>
<td>Propensity x Prior Delinquency</td>
<td>-0.193</td>
<td>0.64</td>
</tr>
<tr>
<td>( \rho (\sigma^2/(\sigma^2+\sigma^2_x)) )</td>
<td>0.076</td>
<td>0.56</td>
</tr>
<tr>
<td>( \alpha )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-531.70</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: \( \rho \) indexes the correlation between the error terms of the various waves for the probit model. The \( \alpha \) and \( \beta \) terms govern the beta distribution from which the negative binomial random effects are drawn.
where $\theta$ is a vector of six probit regression coefficients conformable for multiplication with the vector of six exogenous variables: $q_1 =$ respondent's sex, $q_2 =$ respondent's race, $q_3 =$ whether the family receives public assistance, $q_4 =$ whether the father was unemployed, $q_5 =$ the age of the respondent's youngest parent, and $q_6 =$ the educational level of the respondent's most well-educated parent. In each case, it seems unlikely that variation in any of these variables could be caused by individual levels of criminal propensity or criminality.17

To deal with the problem of zero covariance between the time-varying independent variable and $\tau_i$, we then mean-centered the measure of delinquent peer exposure at each wave. In our estimated equations, then, we included the estimated probability of offending at the first wave from Heckman's approximation and the mean-centered delinquent peer exposure variable. The results of our reanalysis are reported in Tables 5 and 6.

In Table 5, we present the parameter estimates of random-effects probit and negative binomial regression models with the full sample of $N = 471$ cases, and in Table 6 we present the regression models with the sample that had complete data at all four waves of data collection ($N = 390$). These two tables are comparable to Table 3 since they all impose the constraint that prior offending activity and delinquent peer exposure exert similar effects between the high and low antisocial propensity groups.

The results in Tables 5 and 6 suggest that the effect of prior offending activity is substantially attenuated from what was originally reported. We previously reported that the probit coefficient for prior offending was $\gamma = .837$ ($z$ value = 6.84) while holding an attitude-based measure of antisocial propensity constant and $\gamma = .862$ ($z$ value = 7.11) with a behavior-based measure of antisocial propensity held constant (see Table 3). In the corrected reanalysis with complete data (Table 5), these coefficients are reduced to .323 and .363, respectively, and are no longer statistically significant at conventional levels. When cases with missing data are included (Table 6), the attenuation is slightly less, with probit coefficients of .445 and .475 for the attitude and behavior measures of self-control, respectively. Both of these latter coefficients are significant at a .05 level, but

---

17. We would like to emphasize that the Heckman approximation is, in fact, an approximation. By definition, then, it is not a consistent estimator. Because it is not a consistent estimator, the literature provides no clear guidance on exactly how the first wave offending probability should be estimated. We tried several alternative schemes: (1) the method described above; (2) a method in which we included only the covariates in the substantive model (i.e., race, sex, delinquent peer exposure, and an indicator variable for high criminal propensity based on observed individual characteristics); and (3) a method in which we dropped the delinquent peer exposure measure but retained all of the other variables from (2). Of the three methods we compared, the one presented in the text of this paper produced the results at a greatest variance with those presented in our original analysis.
Table 5. Baseline Probit and Negative Binomial Models Using Heckman’s Approximation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Probit Participation Models</th>
<th>Negative Binomial Frequency Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attitude Propensity Measure</td>
<td>Behavior Propensity Measure</td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>z Ratio</td>
</tr>
<tr>
<td>Sex = Male</td>
<td>.522</td>
<td>3.56</td>
</tr>
<tr>
<td>Race = Nonwhite</td>
<td>-.016</td>
<td>0.11</td>
</tr>
<tr>
<td>Time Index</td>
<td>-.120</td>
<td>2.11</td>
</tr>
<tr>
<td>Mean Delinquent Peer Exposure</td>
<td>1.341</td>
<td>7.46</td>
</tr>
<tr>
<td>Δ Delinquent Peer Exposure</td>
<td>.776</td>
<td>4.41</td>
</tr>
<tr>
<td>Prior-Wave Delinquency</td>
<td>.323</td>
<td>1.33</td>
</tr>
<tr>
<td>High-Propensity Indicator</td>
<td>.572</td>
<td>2.81</td>
</tr>
<tr>
<td>$\rho$ ($\sigma^2/\left(\sigma^2+\sigma^2\right)$)</td>
<td>.269</td>
<td>1.83</td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-467.33</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: $\rho$ indexes the correlation between the error terms of the various waves for the probit model. The $\alpha$ and $\beta$ terms govern the beta distribution from which the negative binomial random effects are drawn.
### Table 6. Baseline Probit and Negative Binomial Models Using Heckman’s Approximation with Full Sample ($N = 471$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Probit Participation Models</th>
<th>Negative Binomial Frequency Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attitude Propensity Measure</td>
<td>Behavior Propensity Measure</td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>$z$ Ratio</td>
</tr>
<tr>
<td>Sex = Male</td>
<td>.476</td>
<td>3.65</td>
</tr>
<tr>
<td>Race = Nonwhite</td>
<td>.043</td>
<td>0.34</td>
</tr>
<tr>
<td>Time Index</td>
<td>-.103</td>
<td>1.98</td>
</tr>
<tr>
<td>Any Missing Waves</td>
<td>.199</td>
<td>1.21</td>
</tr>
<tr>
<td>Mean Delinquent Peer Exposure</td>
<td>1.396</td>
<td>8.31</td>
</tr>
<tr>
<td>Δ Delinquent Peer Exposure</td>
<td>.685</td>
<td>4.55</td>
</tr>
<tr>
<td>Prior-Wave Delinquency</td>
<td>.445</td>
<td>2.01</td>
</tr>
<tr>
<td>High-Propensity Indicator</td>
<td>.500</td>
<td>2.79</td>
</tr>
<tr>
<td>$\rho (\sigma^2_1/\sigma^2+(\sigma^2_1+\sigma^2_2))$</td>
<td>.233</td>
<td>1.63</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.631</td>
<td>4.38</td>
</tr>
<tr>
<td>$\beta$</td>
<td>2.770</td>
<td>2.38</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-537.06</td>
<td>-538.29</td>
</tr>
</tbody>
</table>

**NOTE:** $\rho$ indexes the correlation between the error terms of the various waves for the probit model. The $\alpha$ and $\beta$ terms govern the beta distribution from which the negative binomial random effects are drawn.
their magnitude is substantially smaller than what we originally reported. As expected from the simulation evidence presented above, the effect of ignoring initial conditions is to bias positively the estimate of the effect of the lagged outcome variable.\textsuperscript{18}

For the event-count outcomes in Tables 5 and 6, the estimated effects of prior offending activity continue to be positive and statistically significant but their magnitudes are sharply diminished from the values we first reported (see Tables 3 and 4). In our original negative binomial models, we reported that the effect of prior on current offending was .959 and .984 for the attitude and behavior measures, respectively. With the Heckman approximation, however, these coefficients are reduced to .400 (for the complete data) and .500 (when cases with some periods of missing data are included), although they remain statistically significant.

In our original analysis, we reported that the probit coefficient for the effect of delinquent peers on subsequent behavior was approximately .900, and statistically significant. Moreover, an examination of a criminal propensity by delinquent peer exposure interaction indicated that the effect of delinquent peers was comparable for those with high and low criminal propensity.

Tables 5 and 6 report mean-centered revised estimates of the effects of delinquent peer exposure. A key feature of these results is the presence of two estimates of delinquent peer exposure effects. The first estimate represents the effect of individual differences in average levels of delinquent peer exposure over the entire survey period of four years. Our primary interest in the delinquent peer exposure variable, however, is in the effects of year-to-year changes in delinquent peer exposure on offending activity. As Tables 5 and 6 suggest, the estimated effects of delinquent peer exposure are reduced somewhat from the values we originally reported. While shifts in delinquent peer exposure are strongly related to self-reported levels of criminal activity, our conclusion is that the mean-centering procedure of Horney et al. (1995) leads to an (appropriately) more conservative estimate of these effects.

In Tables 7 and 8, we present the parameter estimates associated with an interactive specification. In this analysis, we relax the constraint that delinquent peer exposure and prior offending effects are the same across

\textsuperscript{18} It is important to keep in mind, however, that our simulation evidence indicated that the Heckman approximation provides a negatively biased estimate of the true effect of lagged $y$. Our reported estimate for prior offending with this approximation, then, is a conservative one, and we do not know the precise magnitude of the attenuation. As we have suggested, the advantage of the Heckman approximation is not that it provides an unbiased estimator, but that it provides a less biased estimator than either ignoring the initial-conditions problem completely or applying some other "fix" (such as assigning the first lag $y$ to zero for everyone).
the high and low antisocial propensity groups. In our previous analysis we found that the effects of these variables were similar for both groups. As Tables 7 and 8 suggest, these conclusions continue to hold for the random-effects probit specifications. Our random-effects count data estimator that invokes the behavioral measure of antisocial propensity also leads us to this conclusion. Despite these important similarities, however, the results no longer speak with one voice. When we employ the attitude-based measure of antisocial propensity in the count data estimator, the effect of prior criminal activity on future criminal activity is virtually zero among those with high antisocial propensity, but it is positive and statistically significant for the balance of the population.\textsuperscript{19} Contrary to our original results, this is exactly what typological theorists like Moffitt and Patterson would predict.

**DISCUSSION AND CONCLUSIONS**

Collectively, both our simulation results and our examination of the previously reported findings of Paternoster and Brame (1997) converge to suggest that the initial-conditions problem must be given careful consideration by researchers using panel data. When initial conditions are not adequately resolved, a bias favoring a dynamic explanation of behavior is propagated throughout the model to affect the estimated coefficients on lagged dependent variables, the time-varying and time-stable explanatory variables, and the correlation of the error term over time. In addition, bias is also caused when unobserved heterogeneity is correlated with any time-varying explanatory variables in the model. This bias also favors a dynamic explanation of crime. The bias due to the failure to observe initial conditions appears to be larger than the bias due to the zero-covariance problem. This finding has obvious implications for some of the literature in this area because researchers have employed a lagged measure of offending as a proxy for dynamic causal processes.

The Paternoster and Brame (1997) article discussed above and a similar study by Nagin and Paternoster (1991) used self-reported data from general population surveys and found strong evidence that prior offending was a key correlate of future crime. In contrast, work by Nagin and Farrington (1992b) and Paternoster et. al. (1997) using official data from higher risk samples found little evidence for this conclusion. Paternoster and Brame (1997) speculated that these different results across studies could support a typological theory of crime posited by Moffitt (1993) or

\textsuperscript{19} There was some suggestion of this pattern in our original analysis but it was not statistically significant at conventional significance levels. Importantly, this interaction does not appear to hold with the same strength when the analysis is confined only to cases with valid data for all four waves.
Table 7. Probit and Negative Binomial Models with Interaction Terms Using Heckman's Approximation on Complete Data Sample \((N = 390)\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Probit Participation Models</th>
<th>Negative Binomial Frequency Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attitude Propensity Measure</td>
<td>Behavior Propensity Measure</td>
</tr>
<tr>
<td></td>
<td>Estimate (z) Ratio</td>
<td>Estimate (z) Ratio</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.591 9.36</td>
<td>-3.458 9.91</td>
</tr>
<tr>
<td>Sex = Male</td>
<td>.511  3.53</td>
<td>.535  3.84</td>
</tr>
<tr>
<td>Race = Nonwhite</td>
<td>-.018  0.13</td>
<td>-.011  0.07</td>
</tr>
<tr>
<td>Time Index</td>
<td>-.118  2.05</td>
<td>-.122  2.16</td>
</tr>
<tr>
<td>Mean Delinquent Peer Exposure</td>
<td>1.301  6.98</td>
<td>1.245  7.30</td>
</tr>
<tr>
<td>(\Delta) Delinquent Peer Exposure</td>
<td>.779  4.21</td>
<td>.776  4.22</td>
</tr>
<tr>
<td>Prior-Wave Delinquency</td>
<td>.423  1.68</td>
<td>.319  1.32</td>
</tr>
<tr>
<td>High-Propensity Indicator</td>
<td>.356  0.47</td>
<td>-1.440 1.06</td>
</tr>
<tr>
<td>Interaction Terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propensity x Mean Del. Peer Exp.</td>
<td>.203  0.47</td>
<td>1.020 1.35</td>
</tr>
<tr>
<td>Propensity x (\Delta) Del. Peer Exp.</td>
<td>-.094  0.24</td>
<td>-.075 0.21</td>
</tr>
<tr>
<td>Propensity x Prior Delinquency</td>
<td>-.491  0.81</td>
<td>.279  0.55</td>
</tr>
<tr>
<td>(\rho) ((\sigma^2_0/(\sigma^2_0+\sigma^2)))</td>
<td>.257  1.71</td>
<td>.247  1.71</td>
</tr>
<tr>
<td>(\alpha)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-466.47</td>
<td>-466.56</td>
</tr>
</tbody>
</table>

**NOTE:** \(\rho\) indexes the correlation between the error terms of the various waves for the probit model. The \(\alpha\) and \(\beta\) terms govern the beta distribution from which the negative binomial random effects are drawn.
Table 8. Probit and Negative Binomial Models with Interaction Terms Using Heckman’s Approximation on Full Sample (N = 471)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Probit Participation Models</th>
<th></th>
<th>Negative Binomial Frequency Models</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attitude Propensity Measure</td>
<td>Behavior Propensity Measure</td>
<td>Attitude Propensity Measure</td>
<td>Behavior Propensity Measure</td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>z Ratio</td>
<td>Estimate</td>
<td>z Ratio</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.741</td>
<td>9.94</td>
<td>-3.624</td>
<td>10.61</td>
</tr>
<tr>
<td>Sex = Male</td>
<td>.466</td>
<td>3.62</td>
<td>.484</td>
<td>3.84</td>
</tr>
<tr>
<td>Race = Nonwhite</td>
<td>.044</td>
<td>0.35</td>
<td>.051</td>
<td>0.39</td>
</tr>
<tr>
<td>Time Index</td>
<td>-.099</td>
<td>1.88</td>
<td>-.106</td>
<td>2.03</td>
</tr>
<tr>
<td>Any Missing Waves</td>
<td>.195</td>
<td>1.17</td>
<td>.150</td>
<td>0.89</td>
</tr>
<tr>
<td>Mean Delinquent Peer Exposure</td>
<td>1.379</td>
<td>7.58</td>
<td>1.338</td>
<td>8.15</td>
</tr>
<tr>
<td>Δ Delinquent Peer Exposure</td>
<td>.638</td>
<td>3.95</td>
<td>.677</td>
<td>4.23</td>
</tr>
<tr>
<td>Prior-Wave Delinquency</td>
<td>.564</td>
<td>2.39</td>
<td>.414</td>
<td>1.85</td>
</tr>
<tr>
<td>High-Propensity Indicator</td>
<td>.651</td>
<td>1.05</td>
<td>-.629</td>
<td>0.75</td>
</tr>
<tr>
<td>Interaction Terms</td>
<td>Propensity x Mean Del. Peer Exp.</td>
<td>-.071</td>
<td>0.02</td>
<td>.505</td>
</tr>
<tr>
<td></td>
<td>Propensity x Δ Del. Peer Exp.</td>
<td>.132</td>
<td>0.38</td>
<td>.026</td>
</tr>
<tr>
<td></td>
<td>Propensity x Prior Delinquency</td>
<td>-.484</td>
<td>1.00</td>
<td>.401</td>
</tr>
<tr>
<td></td>
<td>ρ (σ²/(σ²+α²))</td>
<td>.217</td>
<td>1.45</td>
<td>.219</td>
</tr>
<tr>
<td></td>
<td>α</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>β</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Log Likelihood</td>
<td>-535.93</td>
<td>-536.45</td>
<td>-1091.63</td>
</tr>
</tbody>
</table>

NOTE: ρ indexes the correlation between the error terms of the various waves for the probit model. The α and β terms govern the beta distribution from which the negative binomial random effects are drawn.
Patterson (1993), or could be caused by differences between self-report and official data.

Our current analysis now suggests a third possible explanation. Studies with strong results supporting dynamic criminological theories may have overstated the importance of dynamic factors. The reported coefficient estimates for dynamic factors from these studies could be biased because of problems with initial conditions. Both the Paternoster and Brame (1997) and Nagin and Paternoster (1991) studies violate the initial-conditions assumption. These studies are based on self-reported involvement in minor forms of delinquent behavior within a sample of conventional youth, a nontrivial proportion of whom were already involved in offending at the time of the first measurement period. In contrast, the studies that found evidence more supportive of static theories (Nagin and Farrington, 1992a, 1992b; Paternoster et al., 1997) do not violate the initial-conditions assumption. It is important to keep in mind, however, that evidence in this analysis and others clearly continue to suggest that dynamic factors such as delinquent peer exposure are important.

A logical and perhaps immediate response to the estimation problems discussed here might be to avoid specifying models that include lagged offending as an explanatory variable. Since lagged offending has been used as a proxy for other dynamic processes that may have an impact on future offending, and since there is evidence that dynamic processes are important, inclusion of other variables that more accurately reflect the dynamic processes of interest may be more interesting. Time-varying variables other than prior offending may also be valuable from a technical standpoint since the effects of these kinds of covariates can be consistently estimated without appealing to the initial-conditions assumption.

A problem with such an approach is that the failure to include the lagged outcome variable may itself distort matters because there are compelling theoretical reasons to retain it. Nagin and Paternoster (1991) have previously suggested several theoretical routes whereby past offending activity can have an important causal effect on future offending activity. Akers’ (1985) learning theory, for example, predicts that crime can be a positive and rewarding experience for some (netting money, excitement, status from peers), leading to additional crime. Experiencing crime as a positive reinforcement will, other things being equal, lead one to revisit it. A special case of this is Jack Katz’s (1988) notion that crime provides its participants with “sneaky thrills”—thereby seducing them into continued action. The argument that crime has its own rewards is also central to labeling theory’s explanation of secondary deviance (Lemert, 1967).

The problem with the correlation between dynamic factors and unobserved heterogeneity is more easily addressed. Separating the variable into its within-individual and between-individual components is both
straightforward and intuitive. Future work that uses random-effect-like estimation on panel data should begin to report results using this type of decomposition.

We think that another line of future research concerns sources of omitted variable bias in panel analysis not covered by our current work. As we mentioned above, there are three types of omitted variable bias in studies like the kind we discuss here. One type of bias occurs when time-constant omitted variables are correlated with time-constant included variables (type 1). A second type of bias, which was our concern here, occurs when time-constant omitted variables are correlated with time-varying included variables (type 2). Finally, a third type of omitted variable bias occurs when time-varying omitted variables are correlated with time-varying included variables (type 3). Type 1 bias is generic and not unique to panel analysis, and it is very difficult to resolve. Type 3 bias, we think, is more tractable and should be the subject of additional research. For our part, we are continuing to try to resolve these issues and extend an invitation for others to join us.

REFERENCES

Akers, Ronald L.

Allison, Paul D.

Blumstein, Alfred, Jacqueline Cohen, Jeffrey A. Roth. and Christy A. Visher (eds.)

Bryk, Anthony S. and Stephen W. Raudenbush

Bushway, Shawn, Robert Brame, and Raymond Paternoster

Corcoran, Martha and Mary S. Hill

Donovan, John E. and Richard Jessor

Elliott, Delbert S., David Huizinga, and Suzanne S. Ageton
Elliott, Delbert S., David Huizinga, and Scott Menard  
1989 Multiple Problem Youth. New York: Springer-Verlag.

Farrington, David F.  

Fishman, George S.  

Gottfredson, Michael R. and Travis Hirschi  

Greenberg, David F.  

Greene, William H.  

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches  

Heckman, James J.  

Hirschi, Travis  

Hirschi, Travis and Michael R. Gottfredson  

Horney, Julie, D. Wayne Osgood, and Ineke H. Marshall  

Hsiao, Cheng  

Katz, Jack  

Kessler, Ronald C. and David F. Greenberg  

King, Gary  
Land, Kenneth C. and Daniel S. Nagin
1996 Micromodels of criminal careers: A synthesis of the criminal careers and life
course approaches via semiparametric mixed Poisson regression models,
with empirical applications. Journal of Quantitative Criminology

Land, Kenneth C., Patricia L. McCall, and Daniel S. Nagin
1996 Comparison of Poisson, negative binomial, and semiparametric mixed Pois-
son regression models with empirical applications to criminal careers data.
Sociological Methods and Research 24:387–441.

Laub, John H. and Robert J. Sampson
1993 Turning points in the life course: Why change matters to the study of crime.
Criminology 31:301–325.

Laub, John H., Daniel S. Nagin, and Robert J. Sampson
1998 Trajectories of change in criminal offending: Good marriages and the desis-

Lemert, Edwin M.
1967 Human Deviance, Social Problems, and Social Control. Englewood Cliffs,
N.J.: Prentice-Hall.

Liker, Jeffrey K., Sue Augustyniak, and Greg J. Duncan
1985 Panel data and models of change: A comparison of first difference and con-

Loeber, Rolf and Marc LeBlanc
1990 Toward a developmental criminology. In Michael H. Tonry and Norval
Chicago: University of Chicago Press.

Longford, Nicholas T.
1995 Random coefficient models. In Gerhard Arminger, Clifford C. Clogg, and
Michael E. Sobel (eds.), Handbook of Statistical Modeling for the Behav-
ioral and Social Sciences. New York: Plenum.

Maddala, G. S.
1987 Limited dependent variable models using panel data. Journal of Economic
Elgar.

Moffitt, Terrie E.
1993 Adolescence-limited and life-course-persistent antisocial behavior: A

Nagin, Daniel S. and David P. Farrington
1992b The stability of criminal potential from childhood to adulthood. Criminol-

Nagin, Daniel S. and Kenneth C. Land
1993 Age, criminal careers, and population heterogeneity: Specification and esti-

Nagin, Daniel S. and Raymond Paternoster
1991 On the relationship of past and future participation in delinquency. Crimi-

Nagin, Daniel S., David P. Farrington, and Terrie E. Moffitt
1995 Life-course trajectories of different types of offenders. Criminology 33:111-139.

Osgood, D. Wayne, Lloyd D. Johnston, Patrick O'Malley, and Jerald G. Bachman

Paternoster, Raymond and Robert Brame
1997 Multiple routes to delinquency? A test of developmental and general theories of crime. Criminology 35:49-84.

Paternoster, Raymond, Charles W. Dean, Alex Piquero, Paul Mazerolle, and Robert Brame

Patterson, Gerald R. and Karen Yoerger

Petersilia, Joan

Robins, Lee N.

Rowe, David C. and Daniel J. Flannery

Saltzman, Linda S., Raymond Paternoster, Gordon P. Waldo, and Theodore Chiricos

Sampson, Robert J. and John H. Laub

Wilson, James Q. and Richard Herrnstein
Robert Brame is an Assistant Professor in the Department of Criminology and Criminal Justice at the University of Maryland. His interests include the study of criminological theory and applied statistics. Shawn Bushway is an Assistant Professor in the Department of Criminology and Criminal Justice at the University of Maryland. Work on this paper was completed while he was a post-doctoral fellow for the National Consortium on Violence Research at the University of Maryland. His research interests include the relationship between work and crime, statistical methods for use with panel data, desistance, and developmental criminology. Raymond Paternoster is Professor of Criminology at the University of Maryland and a Fellow with the National Consortium on Violence Research. His research interests include the testing of criminological theory and quantitative methods in criminology.