CHAPTER 35

# Estimating Effects over Time for Single and Multiple Units

LAURA DUGAN

## INTRODUCTION

As criminologists, we seek to more thoroughly understand the nature of change. That is, we want to know what leads to change and how we can facilitate improved change, such as less crime and more social welfare. Therefore, most of our research questions inherently ask about the changes over time. We may want to know whether crime subsides after implementing a new policy. Or, we might anticipate that delinquents will behave better once they get a college degree. Yet, much of the published research in our field relies on data measured at only one point of time. For instance, when estimating policy effects, we often compare crime rates in jurisdictions with that policy to those without it. Or, we compare the criminal behavior of college graduates to that of nongraduates. By ignoring the temporal dimension in these research questions, we often confound the cause with the effect. Furthermore, we are especially vulnerable to distorted findings when the model excludes important measures that are also related to our predictors. In fact, LaFree (1999) argues that the downturn in violent crime that began in the early 1990s caught criminologists by surprise because we had relied on cross-sectional information for our predictions.

Perhaps, this issue can be best illustrated by turning to the study of criminal hot spots. Hot spots are spatially concentrated areas of crime that are responsible for a large proportion of the crime activity in an area (Sherman et al. 1989). After observing this dynamic, policing strategies were developed to specifically combat hot spots by concentrating patrols in hot spot areas (Sherman and Weisburd 1995). While studies have consistently identified crime clusters, relatively little research examines how stable hot spots are over an extended period of time (Weisburd et al. 2004). Weisburd et al. (2004) examine the distribution of crime at street segments in Seattle annually for 14 years using trajectory analysis (see Chap. 4, in this volume). They show that hot spots are generally stable over time and that the overall crime trend is driven by changes in a relatively small proportion of places. This research is important because it explicitly models the possible changes in hot spots, more fully informing the police about their dynamics. More recently, others have more directly examined patterns of space–time clustering, demonstrating the importance of integrating the space and time

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_35, 

dimensions to better understand the different spatio-temporal footprints of distinct crimes (Grubesic and Mack 2008), as well as how risk might change once a nearby household is burglarized (Johnson et al. 2007).

While the integration of spatial and temporal analysis is beyond the scope of this chapter, the aforementioned examples raise the importance of connecting our research questions with the appropriate research method. As most of our research questions are inherently dynamic, criminologists have more recently adopted the methods to analyze changes over time. This chapter is designed to introduce the reader to a set of methodological choices to estimate the effects of changes in an independent variable on a dependent variable. I will begin by outlining several methodological options when the scholar has repeated measures of a single unit. I will then discuss several options when many units are repeatedly measured. Finally, this chapter will conclude with a discussion meant to guide the readers' methodological decisions while analyzing dynamic data.

## STRATEGIES TO ESTIMATING EFFECTS OF CHANGES IN X ON CHANGES IN Y

This chapter is not intended to imply that by modeling changes over time we can establish causality. It is, however, intended to demonstrate that if experimental data are unavailable, by modeling temporal data, we can improve our efforts to reduce the model's vulnerability to misleading biases and take us closer to establishing causality. By imposing the temporal restriction that changes in X must have occurred prior to changes in Y, we are better able to rule out the possibility that the reverse relationship is driving any correlation (i.e., Y caused X). For example, if we find in a cross-sectional study that youth who work tend to participate in antisocial behavior, it would be naïve to conclude that working causes delinquency. Instead, those youths who are drawn toward antisocial behavior might choose to work in order to finance their activities. By modeling activity before and after work, the causal relationship can be better disentangled (see Paternoster et al. 2003 for a discussion of employment and delinquency).

Many strategies exist to model temporal data. In order to decide upon the best approach, the criminologist must turn to two sources for guidance, the theory and the data. Ideally, theory can direct the type of data collection that is best suited for each research question. Yet, collecting temporal data takes resources and time, which are often not readily available. Thus, it is often more efficient to rely on existing data sources to address research questions. The remainder of this chapter is organized according to the types of data. I begin by discussing several methodological options available when using data on repeated measures of a single unit. I then turn to the options available when using data on repeated measures of multiple units.

### Repeated Measures of One Unit

Oftentimes, our research question is directed exclusively toward better understanding changes in a macro condition for an important locality, such as a state or country. Perhaps, the most common criminological example is the study of the changes in the U.S. homicide rate (Pridemore et al. 2008; LaFree 2005; LaFree and Drass 2002; McDowall 2002; Kaminski and Marvell 2002) because it has been consistently measured over a long period. Scholars have also used time series to study changes in other patterns, such as burglaries (Chamlin and Cochran 1998), arrest rates (LaFree and Drass 1996) number of fugitives (Goldkamp and Vilcica 2008), and disparities in sentencing outcomes (Stolzenberg and D'Alessio 1994). By tracking these events over time, we can develop a deeper understanding of the conditions that are related to their rise and fall, or boom and bust. In fact, a large portion of research conducted using time series focuses exclusively on changes in the dependent variable (LaFree 2005; Messner et al. 2005; LaFree and Drass 2002; McDowall 2002; O'Brien 1999). Other research models time trends by estimating their changes after a single intervention using interrupted time series (Pridemore et al. 2008; Goldkamp and Vilcica 2008; D'Alessio and Stolzenberg 1995; Cochran et al. 1994). Others incorporate multiple independent variables, while correcting the error structures due to dependence in the measures over time (Cohn and Rotton 1997; LaFree and Drass 1996; Jacobs and Helms 1996). Finally, criminologists have just begun to examine the mutual relationships between multiple time series (Witt and Witte 2000).

While the focus of this chapter is to present methods to estimate the effects of independent variables on a dependent variable over time, I turn first to a brief discussion on important considerations when modeling changes in a dependent variable. Generally, an ideal dependent variable is *stationary*, which means that the error term has a fixed distribution that remains constant over time. Furthermore, in a stationary series, the covariance between any two values is a function of the amount of time between them, not the specific value of time (i.e., $\mathrm{Cov}[y_t, y_s] = \mathrm{f}(t - s)$ not $\mathrm{f}(t)$ or $\mathrm{f}(s)$) (Greene 2008; Banerjee et al. 1993). In essence, we want a dependent variable to appear flat, with a constant variance, without a trend or any evidence of seasonality. To demonstrate the importance of analyzing only stationary series,

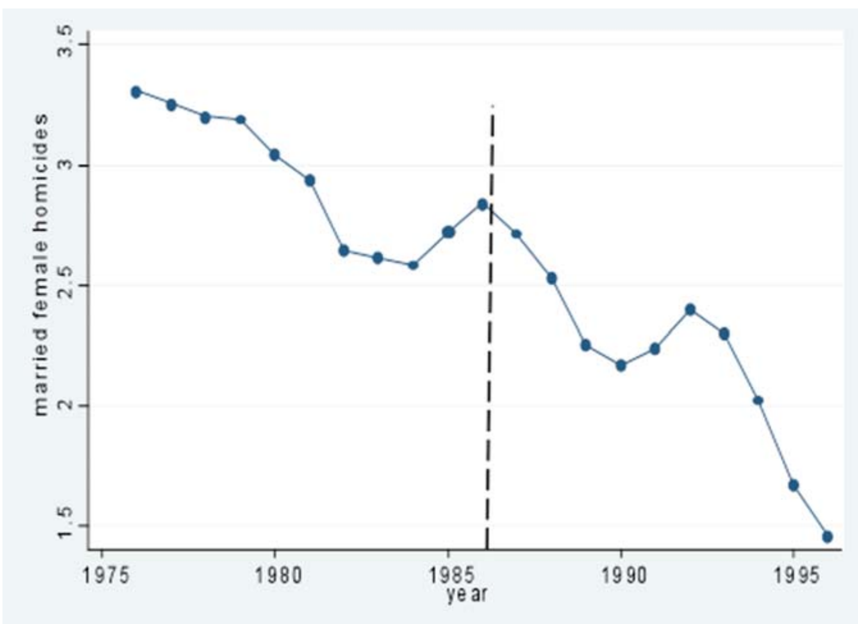


FIGURE 35.1. Nonstationary trend in the married female homicide victimization rate before and after a 1986 intervention. Source: Supplementary Homicide Reports, Uniform Crime Reports.

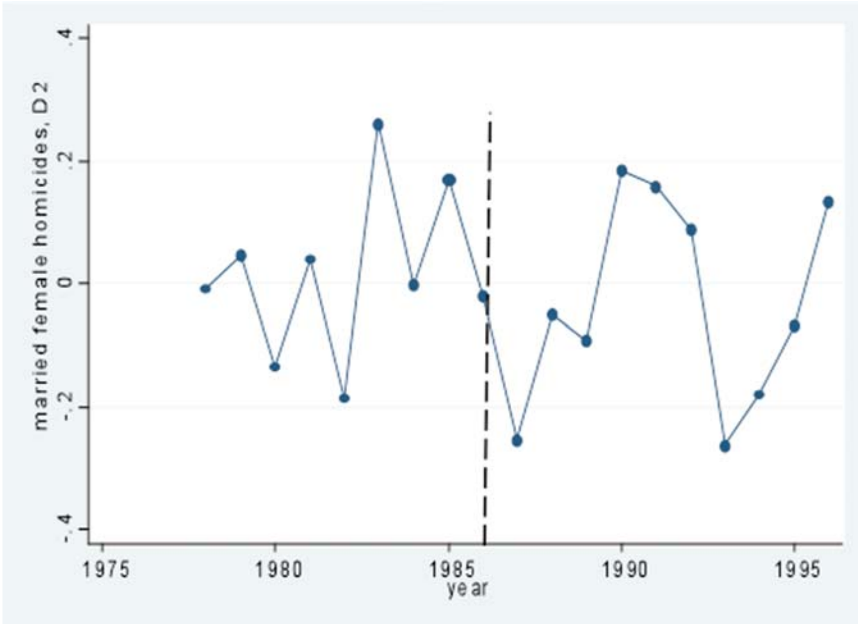


**FIGURE 35.2.** Stationary trend in the married female homicide victimization rate before and after a 1986 intervention. Source: Supplementary Homicide Reports, Uniform Crime Reports.

consider the following example. Figure 35.1 shows the changing rate of homicides perpetrated against wives from 1976 to 1996. The vertical dashed line around 1986 depicts an intervention. If we were to compare the average homicide rates before and after the intervention, we would conclude that homicide was lower after the intervention, suggesting that the intervention successfully reduced homicides. Yet, the drop in homicide after 1986 is part of a larger nonstationary trend that began well before 1986. Fig. 35.2 shows little difference between the average pre- and post-1986 rates, after transforming the data to a stationary series. Had we not addressed the nonstationarity of the series, the findings would have been erroneous.

Tests for stationarity can be preformed, and if the series fails these tests, the data can usually be transformed into a stationary series, depending on the nature of its nonstationarity. For example, if a nonstationary series is found to have a unit root, then by subtracting the previous observation from the current observation (first-differencing) the series will be stationary (Greene 2008). Once the series is stationary, we are better able to predict future observations. More relevant to this chapter, we are able to estimate changes in the series that are due to the changes in other relevant conditions, which could be discrete such as events or policies, or continuous, such as rates or other values.

As noted earlier, value can be gained by exclusively studying patterns of change in dependent variables. O'Brien (1999) examines the arrest rates for both males and females from 1960 to 1995 to determine if the series are converging. He finds that arrest rates are converging for some types of arrests (robbery, burglary, motor theft) but are diverging for homicide. McDowall (2002) examines annual trends of U.S. homicides from 1925 to 1998 to determine if the series is nonlinear. A nonlinear homicide trend suggests that the mechanisms that lead to more homicides are at least partially self-generating (McDowall 2002). He is unable to detect a nonlinear pattern in homicide. LaFree and Drass (2002) test for booms, or rapid increases, in homicide victimization for 34 nations from 1956 to 1998 to determine the validity of the claim by globalizationalists that crime booms are universal since the Second World War. They found little support for that claim.

Since the purpose of this chapter is to discuss the various methods to estimate the effects of independent variables on dependent variables over time, the remainder of the chapter focuses on this type of modeling. First, I will introduce the basic time series model that allows for multiple independent variables, as it corrects for dependence of the error terms. Then, the chapter introduces interrupted time series, which estimates the effects of a single intervention on the time trend. This section of the chapter ends by introducing vector autoregression (VAR) to examine the relationship between two or more dependent variables.

**AUTOREGRESSIVE TIME SERIES.** The simplest of time series models is constructed by simply adding independent variables as predictors of the dependent variable. While simple, since time series data are not generated from a random process, this strategy must consider the inherent dependence across observations that is absent from cross-sectional analysis. The time series process is characterized by its time ordering and the systematic correlation between observations (Greene 2008). Consider the model,

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \tag{35.1}$$

where $\mathrm{cov}(\varepsilon_t,\ \varepsilon_s)$ is a function only of $|t - s|$, the distance between the two observations. This stationary assumption allows us to explicitly model the nature of the correlation between error terms. The covariance is often expressed as the matrix, $\sigma^2\boldsymbol{\Omega}$, where the diagonals of the $\boldsymbol{\Omega}$ are all one and the covariance across observations is depicted by the values of the off-diagonals.

Different types of time series processes will produce different $\boldsymbol{\Omega}$ matrices. Perhaps, the most commonly used is the first-order autoregressive process, AR(1), which has the following error structure,

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \tag{35.2}$$

where $\rho$ is the correlation between all contiguous error terms, and $u_t$ is stationary and independently distributed (i.i.d.). More generally, according to the AR(1) process, $\varepsilon_t$ is a function of $\rho^i\varepsilon_{t-i}$. Thus, for this process, $\rho$ is the only parameter to be estimated, and the effect of prior errors on current errors will fade over time. For example, a relatively large portion of the violent crime rate in January ($\rho$) might directly affect violence in February. Consequently, that smaller portion of February's violence that was affected by January ($\rho^2$) might bleed into March's violent crime rate. This indirect effect is much weaker having gone through February to get to March. By June or July, the consequences of the original violence that has passed from month to month is nearly gone ($\rho^5$ and $\rho^6$).

Furthermore, the autoregressive model becomes more complicated if we think that the actual process includes separate correlations from errors in earlier periods that directly affect the error in the current period. For example, if a delayed shock from two periods ago directly affects the current period (e.g., violence in January directly affects violence in March), we may model the data using an AR(2) model. This would occur if the data had a cyclical or some other more involved pattern.

There are a number of ways to test for first-order autocorrelation. All methods, however, follow the same general idea of estimating the least squares coefficient of $e_t$ on $e_{t-1}$, where $e$ is the residual from the sample data (Greene 2008). Common tests include the Lagrange Multiplier test, Box and Pierce's test (Q-test), and the Durbin–Watson test and are found in most statistical packages.[1] Once the tests are completed and autocorrelation is detected, corrected estimates can be generated from transformations (if $\mathbf{\Omega}$ is know) or by using the Prais–Winsten or Corchrane–Orcutt estimators if $\mathbf{\Omega}$ is unknown (see Greene 2008 for details).[2]

Perhaps, the most cited criminological piece that uses this autoregressive time series approach is work by Cantor and Land (1985), which estimates the effect of levels and changes in unemployment on several types of crimes in the U.S from 1946 to 1982 by hypothesizing that the relationship is a function of both motivation and opportunity. In this article, the authors explore several strategies to stabilize the data by log transforming it and taking first and second differences. They select the model with the fewest number of transformations that produces a Durbin–Watson statistic that indicates no autocorrelation. Findings demonstrate both the positive and negative effects of economic activity on crime. Others have since explored variations to the Cantor and Land study by including inflation rates as an additional measure of economic distress (Devine et al. 1988), by focusing exclusively on youth (Britt 1994), by disaggregating by age and time period (Britt 1997), and by disaggregating by age and race (Smith et al. 1992). In general, the relationship between unemployment and crime is more equivocal among disaggregated groups.[3]

In another analysis, Jacobs and Helms (1996) use time series to estimate the effects of lagged social, political and economic determinants on yearly shifts in U.S. prison admissions from 1950 to 1990. They begin their investigation by testing the series for unit roots and intercorrelations using the augmented Dickey–Fuller test; and find that they need to first-difference the dependent variable and several independent variables in order to stabilize the series. After the series was stationary, they corrected for autocorrelation by using the Marquardt AR1 procedure available in Micro TSP (Jacobs and Helms 1996). Findings suggest that incarceration rates are driven by social inequality, the strength of the Republican Party and whether it was a presidential election year. Surprisingly, unemployment is unrelated to prison admissions. In another analysis, Cohn and Rotton (1997) used 3 hour intervals as their unit of analysis and test whether the relationship between hot temperatures and assaults is stronger during the evening using data on 911 calls and weather conditions in Minneapolis for 2 years. They include in their analysis controls for important temporal indicators, such as the 7 days, of the week, the 12 months of the year, welfare dispersion days and holidays. They calculated Prais–Winsten estimates to correct the standard errors due to the autoregressive error structure of the data. Results confirmed their hypothesis.

Despite the relative simplicity of testing and correcting for the autoregressive process shown in equation (35.1), criminologists have published relatively few articles using this approach. Instead, the literature suggests that we are more drawn toward modeling

[1] If the model includes a lagged dependent variable, the Durbin–Watson statistic will be biased toward no autocorrelation. Turn to Greene (2008: 646) for an adjustment to the Durbin–Watson statistic for the lagged dependent variable. The other two tests produce unbiased estimates regardless of whether the lagged dependent variable is included in the model.

[2] Autocorrelation can also be problematic in panel data. See Greene (2008) for more detail.

[3] This approached used by Cantor and Land (1985) and the others has been criticized by Greenberg (2001). Greenberg explains that nonstationary crime rates are more likely explained by variables that also follow nonstationary processes (unemployment is stationary). Greenberg (2001) shows that while crime and divorce are each nonstationary, when combined they follow a stationary process, suggesting that they are cointegrated (Banerjee et al. 1993). By using error correction models that account for cointegration instead of relying on first- or second-differencing, long term effects can be estimated. See Britt (2001) for continued discussion in this debate.

Autoregressive Integrated Moving Average (ARIMA) models, which more comprehensively address all potential processes of the time series error term. The following section introduces interrupted time series, which is one application of the ARIMA model.

**Interrupted Time Series.** Oftentimes, we are interested in estimating the impact on a series due to a change in a condition or an intervention. For example, D'Alessio and Stolzenberg (1995) test whether the adoption of sentencing guidelines in Minnesota systematically changed the level of incarceration in that state. For this type of problem, the dependent variable is measured repeatedly both before and after the intervention. If the pattern of incarceration after the intervention differs from the preintervention pattern, then results suggest that the intervention had an impact. However, as noted earlier, the inherent dependence of the error terms of time series data will cause estimation problems. Thus, before estimating any intervention effect, we must first adequately model the structure of the error. After accommodating for the error structure using an ARIMA model, differences between the pre- and postintervention series can be tested (McDowall et al. 1980). The three components of this model, autoregressive, integrative, and moving average can be thought of as three specific forms of dependence that can be found in time series data, often referred to as ARIMA(p,d,q).[4] Furthermore, as with the AR(p) process discussed earlier, this dependence can be between the current and previous measures of any order or length of lag. However, in social science, most orders are rather low, most commonly one and sometimes two (Cook and Campbell 1979; McDowall et al. 1980; Greene 2008).

In practice, we begin by determining whether the series is *integrated*, or summed. If the series is summed, it follows a "random walk" pattern, where the next value of $y_t$ is a summation of the starting value of $Y(y_0)$, all the previous random shocks $(\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_{t-1})$, and the current random shock, $\varepsilon_t$. Integrated processes are nonstationary and tend to drift or trend upward or downward depending on the mean of the random shock (McDowall et al. 1980). To formally investigate whether the series is integrated, we can examine the Autocorrelation Function (ACF) which produces the correlations between the current and the lagged values in the series (see McDowall et al. 1980 or Greene 2008 for a derivation of the ACF).[5] If the series is integrated, the correlations will all be large and significant. Integrated series can be made stationary by first-differencing the data, (i.e., $\Delta y_t = y_t - y_{t-1}$). If taking the first differences fails to make the series stationary, the series can be differenced again. In fact, the series should be differenced until the correlations in the ACF are not all large and significant (Greene 2008).[6] In essence, first-differencing removes linear trend, second-differencing removes any quadratic trend, third- removes a cubic trend, and so on (Cantor and Land 1985). The d-value in the ARIMA(p,d,q) model reports the number of times the series is differenced.

[4] Most standard statistical packages run ARIMA models.

[5] $ACF(k) = \dfrac{\sum_{i=1}^{N} (Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N} (Y_i - \bar{Y})^2} \left(\frac{N}{N-k}\right)$ (Cook and Campbell 1979; McDowall et al. 1980).

[6] Some of correlations in the ACF can still be large in a stationary series; however, they will follow a pattern consistent with an autoregressive or moving average process.

Once the series is stationary, it can be examined to determine whether it follows an autoregressive process, a moving average process, or both. The first order autoregressive process, ARIMA(1,0,0),[7] takes the following form:

$$y_t = \Phi_1 y_{t-1} + \varepsilon_t, \tag{35.3}$$

where $\Phi_1$ is the correlation between the current and previous value of $y$. Mathematically, the correlations between the current and earlier lags are ${\Phi_1}^k$, where $k$ is the number of lags (McDowall et al. 1980). Thus, the correlations in the ACF for an ARIMA(1,0,0) process follow the following pattern, $\Phi_1,\ {\Phi_1}^2,\ {\Phi_1}^3, \ldots {\Phi_1}^{t-1}$; such that previous measures have an exponentially decreasing effect on the current value of $y$. Again, the effects of January's violence decreases as they pass through each additional month. When higher orders of the autoregressive process are present, then previous lags have a direct relationship with the current value of $y$ (e.g., January directly affects March or April). More generally, an ARIMA(p,0,0) process takes the following form:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \varepsilon_t. \tag{35.4}$$

As noted earlier, most social science series follow rather low orders of autoregressive processes, remaining usually at an ARIMA(1,0,0) and sometimes an ARIMA(2,0,0) process (McDowall et al. 1980). In fact, Greene (2008) states generally that higher order ARIMA(p,0,0) processes usually indicate that the series is nonstationary.

The first order moving average process, ARIMA(0,0,1), takes the following form:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}, \tag{35.5}$$

where the value of the current $y_t$ is comprised of the current shock, $\varepsilon_t$ and a portion of the preceding shock $\varepsilon_{t-1}$. An important feature of the ARIMA(0,0,1) process is that correlations beyond the first lag drop to zero (McDowall et al. 1980), making it relatively easier to discern between an ARIMA(1,0,0) and an ARIMA(0,0,1) process. The ACF is also a useful tool to identify the order of an ARIMA(0,0,q) process, which is generally written as:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}. \tag{35.6}$$

The ACF for the ARIMA(0,0,q) process will only produce significant correlations for q lags, making it relatively easy to identify the order of the moving average process.

It is more challenging to identify the order of an ARIMA(p,0,0) model. Because the subsequent correlations in the lags of the ACF gradually decrease, an ARIMA(1,0,0) looks quite similar to an ARIMA(2,0,0). Here, the Partial Autocorrelation Function (PACF) can be used. The PACF reports only the direct correlation between the current $y_t$ and $y_{t-k}$. The effects of intervening lags have been "partialed out" of the function (McDowall et al. 1980: 41; Greene 2008). Thus, the PACF for an ARIMA(1,0,0) process only produces a significant correlation for the first lag. That for an ARIMA(2,0,0) process produces significant correlations for the first two lags. In contrast, the correlations in the PACF for an ARIMA(0,0,q)

[7] For simplicity, the ARIMA(1,0,0) is written as if the series needed no differencing. The text could also refer a differenced ARIMA(1,d,0) process.

process gradually reduce to zero with later lags. Thus, the combination of the ACF and PACF allows the scholar to discern between ARIMA(p,0,0) and ARIMA(0,0,q) processes (McDowall et al. 1980).[8] Mixed ARIMA(p,0,q) processes are less common, and are often difficult to discern. In fact, the ACFs and PACFs of ARIMA(2,0,1) and (1,0,2) models look similar (Cook and Campbell 1979).[9] Having said that, there may be one or two spikes in first correlations of the ACF (representing the order of the MA process) followed by more gradual declines in the magnitude of the correlations in the later lags (Greene 2008). McDowall et al. (1980) point out that because of parameter redundancy, mixed ARIMA(p,0,q) processes usually simplify to an ARIMA(p,0,0) or an ARIMA(0,0,q) process.[10]

Time series data can also have seasonal correlations that correspond to systematic and cyclical changes over time. For example, crime generally rises on weekends compared to weekdays (Cohn and Rotton 1997). Thus, in a daily time series of criminal activity, positive correlations are found between observations measured 7 days apart. ARIMA models easily account for seasonal correlations by explicitly estimating autoregressive or moving average parameters for the specified lags. In the daily time series mentioned earlier, the ARIMA model can account for correlations between the current $y_t$ and $y_{t-7}$. See Cook and Campbell (1979), McDowall et al. (1980) or Greene (2008) for a discussion of seasonality.

Once the process has been identified and accounted for, the series is now considered "white noise" or purely random. This allows us to test for a systematic difference between the pre- and postinterruption or intervention period. However, the nature of the intervention effect might differ across contexts. These differences can mostly be characterized in a function that depicts how they begin and how long they lasted (Cook and Campbell 1979; McDowall et al. 1980). Its beginning will be abrupt if the change is immediate once the intervention is implemented. Or the effects of the intervention might be more gradual, slowly or quickly building to its full potential. Furthermore, the intervention might produce permanent or temporary effects. Table 35.1 shows the transfer function, or operationalization, of the intervention used to appropriately estimate the beginning and duration of the effects. Note that $I_{t+}$ is a dummy variable that equal 0 prior to the intervention and 1 after the intervention. Also, $I_t$ is a dummy variable that is 0 for all periods except for that containing the intervention. The parameters ω and δ represent the magnitude of the effect and gradient of its rise or fall, respectively. See Cook and Campbell (1979) for the derivation of these functions.

Notice that no function provides for a gradual, temporary change, which cannot be modeled easily (McDowall et al. 1980). Also notice that the first function is imbedded in the second. Similarly, the second and third are nearly identical. This suggests that the nature of

**TABLE 35.1. Operationalization of estimated effects of an intervention**

| Expected effect | Transfer function |
|---|---|
| Abrupt, constant change | $Y_t = \omega I_{t+} + noise$ |
| Gradual, constant change | $Y_t = \delta Y_{t-1} + \omega I_{t+} + noise$ |
| Abrupt, temporary change | $Y_t = \delta Y_{t-1} + \omega I_t + noise$ |

[8] Most standard statistical packages will produce ACF and PACF tables. For example, the Stata command corrgram will produce a table the autocorrelations, partial autocorrelations, and the Portmantequ (Q) statistics for time series data. The commands AC and PAC produce a graphical version of the ACF and PACF values.

[9] See Cook and Campbell (1979: 247–250) for a more thorough discussion of mixed models.

[10] By repeated substitution, autoregressive processes can take a moving average form (Greene 2008).

the effect can be empirically determined, absent any guiding theory. McDowall et al. (1980) suggest that the scholar start by modeling the abrupt, temporary intervention component. If the estimate of $\delta$ is near one, then the effects of the intervention is likely constant. Next, the scholar should model the gradual, constant change intervention component. If the estimate of $\delta$ is too small and insignificant, then the intervention is best modeled using the function for an abrupt, constant change.

This method has been applied by criminologists to a wide variety of situations. Some have modeled changes in policy, such as work by Loftin et al. (1983) who tested the claim that a 1977 gun law in Michigan that imposed an additional 2 years to a felon's sentence would deter criminal behavior. The only crime that seemed to be affected by the intervention was gun homicide. However, since the number of gun assaults was unaffected by the interventions, the authors conclude that the law was ineffective. Goldkamp and Vilcica (2008) use interrupted time series to find a significant rise in the number of fugitives in Philadelphia after a targeted drug enforcement policy was implemented. Work by Stolzenberg and D'Alessio (1994) and D'Alessio and Stolzenberg (1995) estimates the impact of the adoption of Minnesota's sentencing guidelines on changes in unwarranted disparity and the judicial use of the jail sanction. They find less disparity and more incarceration since implementation. Finally, Simpson et al. (2006) use interrupted time series to detect an increase in arrests after the adoption of a pro-arrest policy for domestic violence.

Criminologists have also used this method to test for changes in outcomes after events that were not imposed by policy. For example, recent work by Pridemore et al. (2008) estimates the effects of the Oklahoma City bombing and the September 11th attacks on homicides at the local and state levels. Their findings show that the attacks neither increased nor decreased lethal violence. Cochran et al. (1994) did find evidence that after Oklahoma executed Charles Troy Coleman in 1990, the number of stranger homicides increased, supporting their brutalization hypothesis. Other nonpolicy relationships that have been tested using interrupted time series include changes in oil prices on burglary (Chamlin and Cochran 1998) and the fall of the Soviet Union on homicide and other social ills (Pridemore et al. 2007).

While interrupted time series implies only one interruption, D'Alessio and Stolzenberg (1995) used an ARIMA model to estimate the impact of the initial adoption of sentencing guidelines and two modifications (42 and 111 months after the initial adoption). Furthermore, other covariates can easily be added to the model as controls. In fact, Autoregressive Moving Average with eXogeneous inputs (ARMAX) refers to ARMA models that allow the error structure to also be estimated by a linear combination of independent variables (Stata Press 2005).

**MULTIPLE TIME SERIES.** Oftentimes, we are interested in how two or more series relate to each other. For example, Witt and Witte (2000) examined the mutual or endogenous relationships between crime, prison, and economic factors, and found among other things, a long-run negative impact of prison on crime. Also, we may want to estimate the impact of one or more exogeneous events on multiple series. When events are exogenous, then their causes are independent of all series outcomes, thus making the causal direction exclusively from the event to the outcomes. Enders and Sandler (1993) estimated the impact of six counter-terrorist interventions on several series of specific terrorist tactical attacks including skyjacking, hostage taking, and others. Their analytical strategy was possible due to research in the early 1970s by Sims (1972) who developed a new approach to examine the causal direction between two dependent variables (money stock and income).[11] With this approach, the inter-dynamics between multiple series of variables avoid relying on the identification assumptions necessary to correctly specify contemporaneous relationships when using structural equation modeling (see Brandt and Williams 2007 for a thorough discussion).

Vector autoregression (VAR) is an interdependent reduced form series of equations where each series is a function of its past values and the past values of the other endogeneous variables.[12] With VAR, the contemporaneous relationships are directly included in the parameterization of the residual covariance (Brandt and Williams 2007). Mathematically, this series of equations is shown as

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11} y_{1,t-1} + \ldots + \beta_{1,p} y_{1,t-p} + \ldots + \beta_{1,m} y_{m,t-1} + \ldots + \beta_{1,mp} y_{m,t-p} + \varepsilon_{1t} \\ y_{2t} &= \beta_{20} + \beta_{21} y_{1,t-1} + \ldots + \beta_{2,p} y_{1,t-p} + \ldots + \beta_{2,m} y_{m,t-1} + \ldots + \beta_{2,mp} y_{m,t-p} + \varepsilon_{2t} \\ &\vdots \\ y_{mt} &= \beta_{m0} + \beta_{m1} y_{1,t-1} + \ldots + \beta_{m,p} y_{1,t-p} + \ldots \beta_{m,m} y_{m,t-1} + \ldots + \beta_{m,mp} y_{m,t-p} + \varepsilon_{mt} \end{aligned}, \tag{35.7}$$

where $m$ is the number of endogeneous variables and $p$ is the number of lags.[13] Brandt and Williams (2007) show that these equations can be simplified by repeating them in their vector form. By allowing $y_t$ to be a vector of all the current measures of all the endogeneous variables, $\boldsymbol{\beta}_\ell$ to be a matrix of all the coefficients, and $c$ to be a vector of intercepts, (35.7) can be simplified to the following form,

$$y_t = c + \sum_{\ell=1}^{p} y_{t-\ell}\boldsymbol{\beta}_\ell + \varepsilon_t. \tag{35.8}$$

Since $\varepsilon_t \sim \mathrm{N}(0, \Sigma \otimes \mathrm{I})$, meaning that there is no autocorrelation between the residuals and their past values or other variables past residuals, the equations can be estimated as a series of separate Ordinary Least Squares models (Brandt and Williams 2007). However, statistical packages like Stata have built to commands to estimate these models and perform related diagnostics (Stata Press 2005).

The above equations naturally lead to the question of how many lags are appropriate for the model. The general consensus is that the lags should be long enough to capture a full cycle of data as well as any residual seasonality. In fact, by modeling the appropriate number of lags, all autocorrelation is absorbed into the model, leaving the residuals as white noise. Following this logic, quarterly data should include about six quarters, and monthly data should include 13 to 15 lags (Brandt and Williams 2007). If the series is short, it may be necessary to limit the number of lags, as Witt and Witte (2000) did, including only two lags for 38 annual observations. Enders and Sandler (1993) used quarterly data and tested their results for lags of 2, 4, and 8. Likelihood ratio tests and information criteria statistics can more formally determine the optimal number of lags appropriate for the model (Brandt and Williams 2007). Since by

[11] He finds that Granger causality is unidirectional from money stock to income, dismissing arguments that changes in income lead to changes in money stock.

[12] This system works better if the variables are stationary. This can be easily determined using the augmented Dickey–Fuller test, which is found in most statistical packages. If the data are nonstationary, it is usually recommended that the analyst estimate Vector Error Correction Models (Stata Press 2005; Greenberg 2001). However, others suggest that there is still value in estimating VAR if the short-term dynamics are the main focus of the analysis (Brandt and Williams 2007).

[13] This notation is borrowed from Brandt and Williams (2007).

using more lags in the model, the residuals are less likely to be autocorrelated (with its own past values) or cross-correlated (with past residuals of other variables), multivariate tests of autocorrelation, such as the Portmanteau test or Legrange multiplier test, can also be used to select the appropriate number of lags. For example, although Witt and Witte (2000) incorporated only two lags, diagnostics show that two lags are sufficient to remove autocorrelation. General practice is to overfit the VAR model with a large number of lags and then to scale back to the most parsimonious model that is absent of autocorrelation (Brandt and Williams 2007). In essence, the lag length should be long enough to produce residuals that are white noise and long enough to produce unbiased and efficient estimates.

Once the lag length is established, it can be determined whether the variables Granger cause one another. Granger causality is described by Brandt and Williams (2007: 32; see also Granger 1969) as "... $Y_t$ Granger causes $Z_t$ if the behavior of past $Y_t$ can better predict the behavior of $Z_t$ than $Z_t$'s past alone." To test for Granger causality, scholars can either conduct a likelihood ratio test or an F-test where the unrestricted model is the full model with all lags from the other exogeneous variables, and the restricted model includes only the lagged dependent variable (Sims 1972; Brandt and Williams 2007).[14] We can also think of this as a test for exogeneity, where exogeneity is the null condition. If, indeed, the results of this test favored the unrestricted model, we would conclude that one or more of the variables are mutually related, or endogeneous. Witt and Witte (2000) were able to show that durable good consumption and female labor force participation were exogeneous to crime.

While it is important to determine whether there are mutual relationships among the variables, exogeneity tests tell us very little about the nature of the relationship between the variables. There are a number of strategies designed to better inform us about the magnitude and dynamics of the estimates generated by the VAR. First, we can decompose the error variance to determine how much of the variance in the predicted path of each variable is due to the past values of the other variables in each equation (Brandt and Williams 2007). In Enders and Sandler's (1993) examination of terrorist tactics, they were able to show that changes in skyjackings explain about 17% of the forecast error variance in attacks against protected persons, such as diplomats and military officials.[15] A second method to estimate the dynamic impact of the variables on one another is to calculate the impulse response function that graphically displays the effect of a shock (usually one standard deviation in magnitude) on itself and other variables. By using confidence bands that incorporate the error generated from the dependent variable as well as that drawn from the covariance between the two variables, we are able to determine the change in the magnitude and significance of all effects over a time horizon (Brandt and Williams 2007).[16] Witt and Witte (2000) present the impulse response functions for their prison-crime model and show, among other things, that the growth in the prison population has a negative impact on crime that disappears by the sixth year.

---

[14] Stata has a post estimation command vargranger that performs pairwise Granger causality tests after running VAR (Stata Press 2005).

[15] This decomposition is done by first inverting the VAR to its Vector Moving Average representation (VMA) and then decomposing the error covariance matrix. This method is sensitive to the ordering of the variables and all orderings should be considered (Brandt and Williams 2007). The Stata post estimation command fevd (forecast-error variance decompositions) calculates these values.

[16] This method also relies on inverting the VAR to its VMA representation and then decomposing the error variance matrix. Once again, all orderings should be considered (Brandt and Williams 2007). The Stata post estimation command irf (impulse-response functions) produce these graphs.

While this section has outlined the value of using VAR to estimate the mutual effects of multiple variables over time, there are limitations to this method. First, all series must be covariance stationary (Brandt and Williams 2007). Nonstationary series, or series with long-memoried processes (unit root or cointegrated) compromise the validity of the Granger causality tests. While solutions to this problem are not as simple as differencing the data, alternate strategies such as fully modified VAR (Freemen et al. 1998) and vector error correction models (VECM) (Brandt and Williams 2007; Greenberg 2001; Banerjee et al. 1993) can be used to estimate the parameters. In fact, Witt and Witte (2000) estimate their model using VECM after an augmented Dickey–Fuller test uncovers cointegration in their series. A second limitation in VAR analysis is that it estimates such a large number of parameters that the estimates could be inefficient and vulnerable to type II error.

**Other Methods to Model Single Units over Time.** Other methods can be used to estimate the effects of an independent variable on a dependent variable for repeated measures of a single unit over time. For example, spline regression can be used to model structural breaks in any stationary time series (see Marsh and Cormier 2002). A breakdate occurs in the period when one of the parameters in the series changes (Hansen 2001). If there is a known intervention or event, interrupted time series can easily be used to model the structural break. However, if the breakdate or dates are unknown, they can be estimated empirically (Marsh and Cormier 2002; Hansen 2001). Messner et al. (2005) use spline regression to locate structural breaks in city-level homicide trends to identify cities that exhibit a meaningful boom and bust cycle.

Another method that has recently been developed to model the effect of events over time is series hazard modeling (Dugan 2009). Standard hazard models estimate the effect of independent variables on the time until *one* event occurs for *many* units. In this extension of the Cox proportional hazard model, the scholar can model the time between *repeated* events for *one* unit. This strategy was first used in research by Dugan et al. (2005) to estimate the impact of several policies on the hazard of an aerial hijacking by modeling the time until the next hijacking as a function of the current policy profile, and a set of characteristics of the current hijacking event. LaFree et al. (2009) later use this strategy to estimate the impact of several operations implemented by the British government on terrorism by Republicans in Northern Ireland. Most recently, Dugan et al. (2009) use series hazard modeling to examine the effects of a devastating terrorist attack on the continued activity of two Armenian terrorist organizations.

**Limitations.** While time series methods improve our efforts to understand dynamic relationships between variables, they do have their limits. First, because time series models the activity of only one unit, findings are rarely generalizeable beyond that unit. This is especially problematic with smaller units, such as neighborhoods, cities, and states. Generalizeabily is less of an issue when the unit includes the entire country or world. However, modeling large units has its own problems. Macro dynamics tell us very little about the variation across subunits. For example, models that estimate the impact of economic stressors on crime in the U.S. will likely fail to adequately describe how crime responds to economic depression in Atlanta, despite Atlanta's contribution to the data. The overall U.S. trend likely fails to uniformly describe the trends in all U.S. cities, counties, and states. Recall the finding in Weisburd et al. (2004) reporting that macro trends in Seattle were driven by activity in only a few neighborhoods. Finally, a third limitation of time series analysis is that it requires long periods of time in order to efficiently estimate models. Yet, it is often difficult to collect standardized measures over an adequately long period. One strategy to increase the degrees of freedom is to shorten the time of measurement (e.g., from years to quarters or months). However, some measures might be unavailable for the shorter periods. Furthermore, even if the data were available, the pattern in a more refined series might be too sporadic to adequately model.

## Repeated Measures of Multiple Units

One strategy to improve upon all of the limitations mentioned above is to model an outcome using repeated measures of multiple units. For example, instead of modeling data in a series for a single state, we can model data from all 50 states, or a subset of states. More locations can be included in the model to improve generalizabilty and efficiency. Once the scholar has decided to use repeated measures of multiple units, the data need to be properly structured for longitudinal analysis. In contrast to cross-sectional data where each row represents one unit (N), for longitudinal data analysis, each row represents one unit (N) for one period of time (T) (e.g., a city-month or a person-year). Thus, instead of having N rows of data, we now have N $\times$ T rows.

The next decision is how to analyze the data. One simple method would be to run the analysis on the pooled N $\times$ T observations as if the data were cross-sectional as shown in the following equation,

$$y_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + \mathbf{Z}'_{i}\boldsymbol{\alpha} + \varepsilon_{it}, \tag{35.9}$$

where $y_{it}$ is the value of the dependent variable for unit $i$ during period $t$.[17] $\mathbf{X}'_{it}$ is a vector of the time varying independent variables for that same unit and time period and $\mathbf{Z}'_{i}$ is a vector of time invariant independent variables for unit $i$ (note that $\mathbf{Z}$ has no subscript $t$). Here, some variables in $\mathbf{Z}$ might be observed and others may be unobserved. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are the corresponding vectors of parameter coefficients for $\mathbf{X}$ and $\mathbf{Z}$; and $\varepsilon_{it}$ is the error term for unit $i$ during period $t$.[18] While this model is rather simple, it will produce misleading results because it ignores the inherent dependence across the rows of data. The value of the error term, $\varepsilon_{it}$, for the same unit will be correlated across time, and error terms for the same time period will be correlated across units. Mathematically, $\mathrm{Cov}(\varepsilon_{it}, \varepsilon_{it}) \neq 0$ for $i = i$ and $t \neq t$ or for $i \neq i$ and $t = t$. By ignoring this dependence, standard errors will be artificially deflated, possibly producing erroneous significance (Greene 2008). Furthermore, if any of the unobserved independent variables in $\mathbf{Z}'_{i}$ is correlated with any of the observed variables in $\mathbf{X}'_{it}$ or $\mathbf{Z}'_{i}$, then the estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ will be biased (Hausman and Taylor 1981).

This section formally presents two strategies to addressing the dependence of the error terms in panel data: fixed effects and random effects models. Before choosing the method that is most suitable to the research question, scholar must carefully consider the context of analyses. For example, since the fixed effects model absorbs the variation across units (population heterogeneity), theoretical approaches that make predictions that rely on this variation might be better modeled with random effects. Conversely, if important time-invariant variables are

[17] The notation for all equations is borrowed from Greene (2008).

[18] We could easily add another term, $\mathbf{T}'_{t}\boldsymbol{\gamma}$, to represent those variables that change over time, but are stable across units.

missing from the model and we suspect that these omitted variables are correlated with those in the model, the fixed effects model might be the better choice. These considerations are elaborated upon below.

**FIXED EFFECTS LONGITUDINAL ANALYSIS.** Consider the following variation on equation (35.9),

$$y_{it} = \mathbf{X'}_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}. \tag{35.10}$$

In this model, $\mathbf{Z'}_i\boldsymbol{\alpha}$ is replaced by $\alpha_i$, which is an individual specific constant term.[19] By including these "fixed effects," we are essentially controlling for all variation across units. Thus, the model is estimated based on the variation in changes within each individual ignoring variation across individuals. In practice, this term is calculated by either producing dummy variables for all but one individual (N − 1) or by subtracting the unit-specific means from all values of $\mathbf{X'}_{it}$ (most statistical packages are now able to automatically run fixed effects models). The fixed effect thus provides a unit-specific intercept vector, $\alpha_i$, which estimates the deviation of that unit from the overall intercept. Since all of the time-invariant observed and unobserved variables are controlled for, the coefficient estimates are now estimating the effects of changes in **X** on changes in **Y**.

While fixed effects models can be used with any level of analysis, most criminological applications tend to apply it to spatial units rather than to individuals.[20] For example, research by Dugan et al. (1999) combines data from 29 cities over multiple years to produce four time periods between 1976 and 1992 in order to estimate the effects of changes in domestic violence resources on changes in rates of intimate partner homicide (see Dugan et al. 2003 for a more recent analysis). Their model adds a fixed effect for each of the 29 cities ($\alpha_i$) and a fixed effect for each of the four time periods, effectively removing the common average from every city and wave in the data. It was especially important to add time fixed effects because one could easily imagine many unobserved changes between 1976 and 1992 that are correlated with both domestic violence resources and intimate partner homicide. To reduce any bias due to simultaneity, most of the variables are lagged by one wave. Thus, the estimates are unaffected by the adoption of policies due to salient homicides during the same reporting period. Other criminological research that incorporates fixed effect models examine the effects of state-level unemployment on crime (Levitt 2001; Raphael and Winter-Ebmer 2001), and of individual-level employment on delinquency (Paternoster et al. 2003; Apel and colleagues 2008) and on victimization (Ousey et al. 2008).

Despite the importance of controlling for unobserved variation across units, fixed effects panel models have noteworthy limitations that may preclude scholars from relying exclusively on this analytical strategy. First, by controlling for each unit, the model loses efficiency, possibly dropping statistical power below levels needed to detect significance. This issue is especially problematic for panels with large N, since each fixed effect reduces the model by one degree of freedom. Relatedly, by absorbing the variation across units, little variation is left in the model from which the estimate can be drawn (Levitt 2001). Dugan et al. (2003) addressed this limitation in their more recent city-level analysis of domestic violence resources on intimate partner homicide by estimating the models using no-place effects,

---

[19] Note that we could also add a time specific constant that captures variation that is constant across units, but varies over time.

[20] Exceptions are found in the more recent population heterogeneity versus state dependency literature, which will be discussed below.

city-effects, and state-effects. Since the city-level effects were expected to be suffering least from omitted variable bias, those estimates were used to create upper and lower confidence bounds to test for possible bias in the state and no-place effects models. When the estimates from the no-place model fell within the confidence bounds, that estimate was chosen over that from the state-effects model.

Another limitation is that fixed effects modelers may become overconfident that all omitted variable bias is accounted for by the fixed effects. This model only accounts for time invariant (or unit invariant) omitted variables. It fails to account for unmeasured changes in omitted variables that might be correlated with changes in one or more of the included independent variables. Apel and colleagues (2008) directly address this issue in their study of the effects of youth employment on deviant behavior and academic achievement by using child labor laws as an instrument for within unit changes in the youths' work behavior in the fixed effects model (called fixed effects IV). Results show that the fixed effects IV models are better able to isolate the relationship between work intensity and six different outcomes (Apel and colleagues 2008). IV estimation was also used by Ousey et al. (2008) in their investigation of state dependence in victimizations because the fixed effects transformation produced negative biases in the lagged dependent variable.[21]

The above weaknesses are all methodological and should be carefully considered within the context of all panel research designs. However, at least one important substantive concern should also be carefully considered. Theoretical motivations that rely on variation across individuals cannot be tested using the fixed effects model, since that variation is essentially removed from the model. One prominent example in our literature is the debate on whether the correlation between past and current criminal behavior is more dependent on population heterogeneity, meaning differences across individuals, or on state dependence, which argues that past offending impacts subsequent criminal behavior through its influence on other variables that raise the need or desire to offend (see Nagin and Paternoster 1991, 2000). Fixed effects models essentially remove all population heterogeneity from the data, thus only informing the state dependent side of the debate. Those determined to explicitly estimate the effects of population heterogeneity have turned to random effects models.

**RANDOM EFFECTS LONGITUDINAL ANALYSIS.** If we can reasonably assume that the unobserved variables, $\mathbf{Z}'_i$, are uncorrelated with those included in the model, then modeling with random effects is a sensible choice. To further understand the random effects model, the following equation expands the $\mathbf{Z}'_i\boldsymbol{\alpha}$ term in (35.9) by adding and subtracting $\mathrm{E}\left[\mathbf{Z}'_i\boldsymbol{\alpha}\right]$ (Greene 2008).

$$y_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + \mathrm{E}\left[\mathbf{Z}'_i\boldsymbol{\alpha}\right] + \left\{\mathbf{Z}'\boldsymbol{\alpha} - \mathrm{E}\left[\mathbf{Z}'_i\boldsymbol{\alpha}\right]\right\} + \varepsilon_{it} \tag{35.11}$$

Since $\mathrm{E}\left[\mathbf{Z}'_i\boldsymbol{\alpha}\right]$ is a constant, it can be replaced by $\alpha$, and interpreted as the mean of the unobserved heterogeneity. The third term shows the deviation of $\mathbf{Z}'_i\boldsymbol{\alpha}$ from its expected value, which is an error term for the cross-sectional observations. By replacing that deviation with $u_i$, we get the following equation,

$$y_{it} = \mathbf{X}'_{\mathbf{it}}\boldsymbol{\beta} + \alpha + u_i + \varepsilon_{it}. \tag{35.12}$$

[21] Their method differs from that used by Apel and colleagues (2008). Here they incorporated IV estimated with a generalized-method of moments framework.

This equation looks quite similar to the standard regression equation, except that the error term is partitioned into a group specific random element, $u_i$, and the standard $\varepsilon_{it}$, which is independently and identically distributed normally with a zero mean. We can think of the $u_i$ as the error term that would naturally result from randomly selecting the cross-sectional observations from a larger population (Greene 2008). By modeling with random effects, the number of parameters to be estimated is greatly reduced, compared to the fixed effects model. However, as with standard regression analysis, the disturbance, $u_i$ is assumed to be normal, an assumption that may not always be reasonable (Maddala 1987; Nagin and Paternoster 1991; Bushway et al. 1999).

As noted earlier, when criminologists use individuals as the unit of analysis (rather than geographic units), we tend to apply random effects models more often than fixed effects models. This distinction makes sense because it is more reasonable to assume that individuals are randomly drawn from a larger population than are cities, states, or countries. For example, as an introduction to the debate noted earlier, Nagin and Paternoster (1991) use random effects to discern whether the observed relationship between past and current offending is more driven by unobserved heterogeneity or state dependence. They were the first in our field to model their data using random effects. Their findings strongly suggested that persistent offending is mostly driven by state dependence, which contrasted largely with the common belief at that time that persistence in offending is largely due to population heterogeneity (Nagin and Paternoster 1991). These scholars continued with other colleagues to model individuals using random effects and consistently found evidence that state dependence explains, in part, the continuity in an offending career (Paternoster et al. 1997; Paternoster and Brame 1997; Nagin and Farrington 1992a, b).

As with fixed effects models, the random effects approach has its limitations. Perhaps most important, is that it relies on two very strong assumptions. First, it assumes that the unobserved heterogeneity is unrelated to the covariates in the model (Greene 2008). If $u_i$ is correlated with $\mathbf{X}'_{it}$, then estimates of $\boldsymbol{\beta}$ will be biased and inconsistent (Hausman and Taylor 1981). This is a strong assumption that modelers face, even with cross-sectional regression analysis, where we are rarely certain that the error term is exogenous to the independent variables. Yet, Hausman and Taylor (1981) argue that by following the cross-section over time, as in panel analysis, we can test for this independence (see also Hausman 1978; Wu 1973; and Durbin 1954). Furthermore, they introduce an instrumental variable strategy that identifies an uncorrelated portion of $\mathbf{X}'_{it}$ to produce a consistent estimate of $\boldsymbol{\beta}$ (Hausman and Taylor 1981). The key is to use an "instrument" that is correlated with $\mathbf{X}'_{it}$ and $\mathbf{Y}$, but is uncorrelated with omitted variables to isolate the portion of $\mathbf{X}'_{it}$ that is invulnerable to omitted variable bias. Note that this is a similar strategy adopted by Apel and colleagues (2008) to address dependence in the fixed effects model. While finding an adequate instrument will better isolate the "true" effects, such instruments are often very difficult to find, requiring creativity, available data, and much luck. Hence, only a handful of criminological studies have successfully used instrumental variables to reduce omitted variable bias; and the remaining studies cautiously draw conclusions acknowledging their vulnerability to bias.

The second assumption in random effects models is that the $u_i$ term is independently and identically normally distributed (Nagin and Paternoster 1991). Nagin and Paternoster (1991) warn that the results can be very sensitive to this distributional assumption, leading to biased estimates. Bushway et al. (1999) directly address this issue by modeling the same research question (unobserved heterogeneity versus state dependence) using different approaches to see if the results would lead to different conclusions. They compare the estimates generated from a random effects model, a fixed effects model and a semiparametric model (developed in Nagin and Land 1993) of the effect of prior offending on current offending in the 1958 Philadelphia cohort (Tracy et al. 1990). Their findings show that all three methods produce a similar, and therefore robust, estimate of the state dependence parameter (Bushway et al. 1999). Others have similarly compared findings from random and fixed effects models and found that the estimate of within variation for each produces similar results (Paternoster et al. 2003; Phillips and Greenberg 2008).

In summary, when deciding whether to use fixed or random effects, it is necessary to understand how the strengths and weaknesses of each will affect the modeling context. If the covariates in the model are likely to be correlated with the unmeasured heterogeneity, then a fixed effect model may be the best choice. However, this model is problematic if important covariates are constant over time, for they will be washed out by the fixed effects. Also, this model uses a large number of degrees of freedom, making it inefficient, possibly leading to type II errors. Another option is to use the random effects model and address the dependence of covariates on the unobserved heterogeneity by applying IV estimation to the model, as demonstrated by Hausman and Taylor (1981). To be sure that the findings of the random effects model are robust to the distributional assumption, one might follow the lead of other researchers by comparing their estimate of within subject variation to that generated from the fixed effects model. However, caution must be made when estimating the effect of a lagged dependent variable, because estimates from both the random and fixed effects models may produce erroneous results (Ousey et al. 2008).

**OTHER LONGITUDINAL METHODS.** Fixed and random effects are by no means the only methods to model panel or longitudinal data. Multilevel, or random coefficient, modeling strategies have also been used to estimate effects using this data. One advantage of multilevel modeling is that it allows linear trends to vary across units (Bryk and Raudenbush 1992).[22] Rosenfeld et al. (2007) use multilevel modeling (hierarchical linear models) to estimate the effect of order-maintenance arresting on robbery and homicide rates for New York City precincts; and find modest effects. Two other methods to model longitudinal data are described in other chapters in this volume. Group-based modeling, or trajectory analysis, was developed by Daniel Nagin as a method to estimate the developmental trajectories of individuals over the life course (Chap. 4 – Nagin 2009). Independent variables can be added to the model as predictors of group membership (Nagin 2005). A similar methodology, latent growth curve trajectories can also be used to model the changes in individuals' developmental trajectories. Petras and Masyn (2009) demonstrates in this book how to analyze longitudinal data using this method in Chap. 5 of this book.

## DISCUSSION

Because so much of our theoretical and practical research interests in criminology typically follow the question "If X changed, would Y change too?" this chapter was designed to introduce the basic foundations of estimating the effects of X on Y for time series and longitudinal data. While the universe of modeling options for both of these data types encompasses far

[22] Multilevel modeling can be accomplished using a variety of statistical software packages such as Stata, SAS and SPSS. The packages HLM, MLwinN, aML, and WINBUGS were explicitly designed to estimate multilevel models.

more than what could reasonably be included in a single chapter, readers should by now have a reasonable idea of which methods are most appropriate for their research questions, given the available data.

When either the research question or the available data limits the scholar to modeling one unit over time, she or he could choose from several methods of time series. Regardless of methodological choice, all seem to follow the general theme of first assuring that the series is stationary, and then modeling the correlation of the error structure. From the literature, it appears that the decision to use an autoregressive model or an ARIMA model is based on individual preference and discipline. As criminologists, we have used both, while more recently preferring ARIMA. Tests for autocorrelation may not detect moving average processes, making ARIMA modeling the more comprehensive choice. Also, by examining the output from the ACF and PACF functions, we can easily observe the error structure of the uncorrected and corrected models making our modeling decisions more tangible. Further, by including an intervention, interrupted time series allows us to more intuitively estimate the changes in temporal patterns after an event. Finally, since ARIMA models can easily incorporate multiple independent variables, there are fewer advantages to relying only on autoregressive time series.

Finally, we have become increasingly aware of how important it is to account for the intricate relationships between multiple series. Often the causal relationship we hope to model is confounded by mutual dynamics among the predictors. For example, social conditions could be driven by criminal activity which, in turn, might lead to more or less crime, shifting social conditions, and so on. While researchers have found that unemployment is exogenous to crime, it has been shown that divorce rates follow the same patterns as crime rates, suggesting that both series should be modeled together (Greenberg 2001). Also, exogenous variables could affect a series either directly or indirectly through a related series. By modeling multiple series, we can disentangle these relationships and more precisely understand the nature of how X affects Y. Multiple series can easily be modeled using VAR for stationary series and error correction models for nonstationary or cointegrated series.

When the research question is best answered with repeated measures of multiple units, the scholar has several methodological choices. Figure 35.3 maps out a subset of these choices using a decision tree. The bottom of this figure shows only four strategies to modeling longitudinal data (the non-italicized referring to the models discussed in this chapter). As demonstrated by other chapters in this volume, there are more than four ways to model longitudinal data.

The first question raised by this map is *whether the variation across units is theoretically important*. This usually can be answered by whether or not the theory requires the researcher to estimate the effects of time-stable variables. In other words, if the only source of variation for an important theoretical variable is across units, then the answer to this question is likely *yes*. Once it is decided that variation across units is important, then the researcher must decide *how vulnerable the model is to omitted variable bias*. If there is *no* obvious vulnerability, then the researcher could use a *random effects model*. If the model is, *indeed*, vulnerable to such bias (i.e., there are important measures that are likely correlated with both the included variables and the dependent variable), then the researcher should attempt to account for possible bias by incorporating methods such as *IV estimation*.

If the answer to the first question is *no*, meaning that there is *no theoretical reason to account for variation across units*, then the scholar must decide whether it is important to account for *differences in the slopes for each unit*. If *yes*, then *multilevel modeling* might be

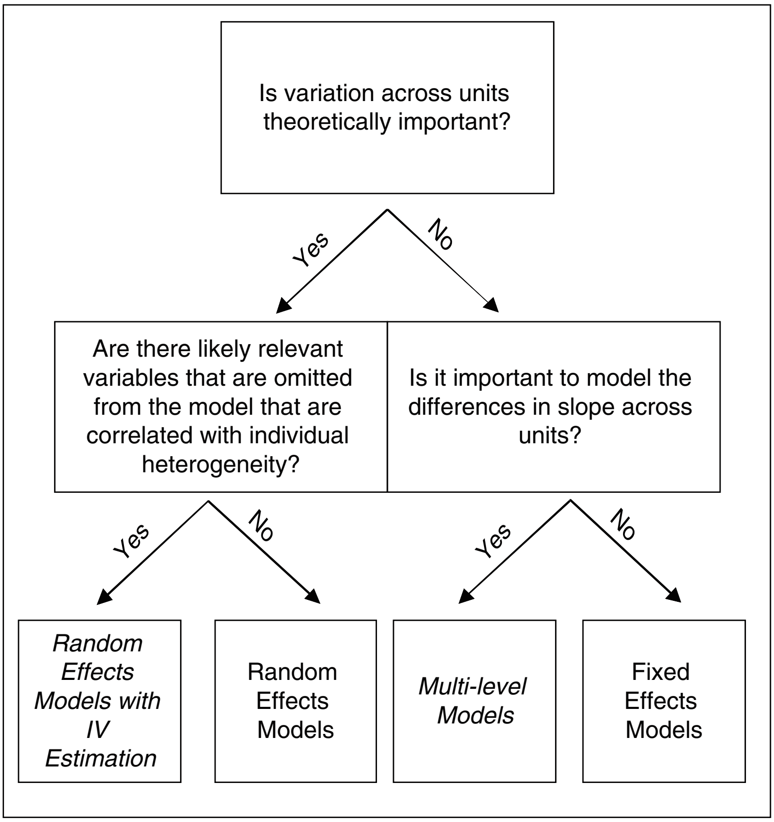


FIGURE 35.3. Decision tree for selecting longitudinal model.

the more efficient choice. While fixed effects models can estimate slopes by including interactions, this specification will require many more degrees of freedom, making the results vulnerable to type II errors. If there is *no* theoretical or practical justification for modeling the slopes for the individual units, then the *fixed effects* model may be the best choice. Admittedly, this decision tree is unrealistically simplistic, as it fails to account for other important issues that might lead the researcher to another decision, such as low statistical power. It does, however, provide a starting point for exploring methodological options.

## REFERENCES

Apel R, Bushway SD, Paternoster R, Brame R, Sweeten G (2008) Using state child labor laws to identify the causal effect of youth employment on deviant behavior and academic achievement. J Quant Criminol 24(4):337–362

Banerjee A, Dolado JJ, Galbraith JW, Hendry DF (1993) Co-integration, error correction, and the econometric analysis of non-stationary data. Oxford University Press, Oxford

Brandt PT, Williams JT (2007) Multiple time series models. Sage Publications, Inc., Thousand Oaks, CA

Britt CL (2001) Testing theory and the analysis of time series data. J Quant Criminol 17(4):343–357

Britt CL (1997) Reconsidering the unemployment and crime relationship: variation by age group and historical period. J Quant Criminol 13(4):405–428

Britt CL (1994) Crime and unemployment among youths in the United States, 1958–1990: a time series analysis. Am J Econ Sociol 53(1):99–109

Bryk AS, Raudenbush SW (1992) Hierarchical linear models. Sage Publications, Inc., Newbury Park, CA

Bushway S, Brame R, Paternoster R (1999) Assessing stability and change in criminal offending: a comparison of random effects, semiparametric, and fixed effects modeling strategies. J Quant Criminol 15(1):23–61

Cantor D, Land KC (1985) Unemployment and crime rates in the post-World War II United States: a theoretical and empirical analysis. Am Sociol Rev 50:317–332

Chamlin MB, Cochran JK (1998) Causality, economic conditions, and burglary. Criminology 36(2):425–440

Cochran JK, Chamlin MB, Seth M (1994) Deterrence or brutalization? An impact assessment of Oklahoma's return to capital punishment. Criminology 32(1):107–134

Cohn EG, Rotton J (1997) Assault as a function of time and temperature: a moderator-variable time-series analysis. J Pers Soc Psychol 72(6):1322–1334

Cook TD, Campbell DT (1979) Quasi-experimentation design & analysis issues for field settings. Houghton Mifflin Company, Boston

D'Alessio SJ, Stolzenberg L (1995) The impact of sentencing guidelines on jail incarceration in Minnesota. Criminology 33(2):283–302

Devine JA, Sheley JF, Smith MD (1988) Macroeconomic and social-control-policy influences on crime rate changes, 1948–1985. Am Sociol Rev 53:407–420

Dugan L (2009) Series hazard modeling: an extension to the cox proportional hazard model to estimate intervention effects on recurrent events for a single unit. Unpublished manuscript. The University of Maryland

Dugan L, Huang J, LaFree G, McCauley C (2009) The Armenian secret army for the liberation of Armenia and the justice commandos of the Armenian genocide. Dynamics of Asymmetric Conflict (forthcoming)

Dugan L, LaFree G, Piquero A (2005) Testing a rational choice model of airline hijackings. Criminology 43(4): 1031–1066

Dugan L, Nagin D, Rosenfeld R (2003) Exposure reduction or retaliation? The effects of domestic violence resources on intimate partner homicide. Law Soc Rev 27(1):169–198

Dugan L, Nagin D, Rosenfeld R (1999) Explaining the decline in intimate partner homicide: the effects of changing domesticity, women's status, and domestic violence resources. Homicide Stud 3(3):187–214

Durbin J (1954) Errors in variables. Rev Int Stat Inst 22:23–32

Enders W, Sandler T (1993) The effectiveness of antiterrorism policies: a vector-autoregression-intervention analysis. Am Polit Sci Rev 87(4):829–844

Freemen J, Houser D, Kellstedt PM, Williams JT (1998) Long-memoried processes, unit roots, and causal inference in political science. Am J Pol Sci 42(4):1289–1327

Goldkamp JS, Vilcica ER (2008) Targeted enforcement and adverse system side effects: the generation of fugitives in Philadelphia. Criminology 46(2):371–409

Granger CWJ (1969) Investigating causal relations by economic models and cross-spectral methods. Econometrica 37(3):424–438

Greenberg DF (2001) Time series analysis of crime rates. J Quant Criminol 17(4):291–327

Greene WH (2008) Econometric analysis, 6th edn. Pearson Prentice Hall, Upper Saddle River, NJ

Grubesic TH, Mack EA (2008) Spatio-temporal interaction of urban crime. J Quant Criminol 24(3):285–306

Hansen BE (2001) The new econometrics of structural change: dating breaks in U.S. labor productivity. J Econ Perspect 15(4):117–128

Hausman J (1978) Specification tests in econometrics. Econometrica 46:1251–1271

Hausman JA, Taylor WE (1981) Panel data and unobservable individual effects. Econometrica 49(6):1377–1398

Jacobs D, Helms RE (1996) Toward a political model of incarceration: a time-series examination of multiple explanations for prison admission rates. Am J Sociol 102(2):323–357

Johnson SD, Bernasco W, Bowers KJ, Elffers H, Ratcliffe J, Rengert G, Townsley M (2007) Space-time patterns of risk: a cross national assessment of residential burglary victimization. J Quant Criminol 23(3):201–219

Kaminski RJ, Marvell TB (2002) A comparison of changes in police and general homicides: 1930–1998. Criminology 40(1):171–190

LaFree G (2005) Evidence for elite convergence in cross-national homicide victimization trends, 1956–2000. Sociol Q 45:191–211

LaFree G (1999) Declining violent crime rates in the 1990s: predicting crime booms and busts. Annu Rev Sociol 25(1):145–68

LaFree G, Drass KA (2002) Counting crime booms among nations: evidence for homicide victimization rates, 1956 to 1998. Criminology 40(4):769–800

LaFree G, Drass KA (1996) The effect of changes in intraracial income inequality and educational attainment on changes in arrest rates for African Americans and whites. Am Sociol Rev 61(3):614–634

LaFree G, Dugan L, Korte R (2009) Is counter terrorism counterproductive? Northern Ireland 1969–1992. Criminology 47(1):501–530

Levitt SD (2001) Alternative strategies for identifying the link between unemployment and crime. J Quant Criminol 17(4):377–390

Loftin C, Heumann M, McDowall D (1983) Mandatory sentencing and firearms violence: evaluating an alternative to gun control. Law Soc Rev 17(2):287–318

Maddala GS (1987) Limited dependent variable models using panel data. J Hum Resour 22(3):307–338

Marsh LC, Cormier DR (2002) Spline regression models. Sage Publications, Inc., Thousand Oaks, CA

McDowall D (2002) Tests of nonlinear dynamics in U.S. homicide time series, and their implications. Criminology 40(3):711–735

McDowall D, McCleary R, Meidinger EE, Hay RA (1980) Interrupted time series analysis. Sage Publications, Inc., Thousand Oaks, CA

Messner SF, Deane GD, Anselin L, Pearson-Nelson B (2005) Locating the vanguard in rising and falling homicide rates across U. S. cities. Criminology 43:661–696

Nagin DS (2009) Group-based modeling: an overview. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York

Nagin DS (2005) Group-based modeling of development. Harvard University Press, Cambridge, MA

Nagin DS, Farrington DP (1992a) The stability of criminal potential from childhood to adulthood. Criminology 30(2):235–260

Nagin DS, Farrington DP (1992b) The onset of persistence of offending. Criminology 30(4) 501–523

Nagin D, Land KC (1993) Age, criminal careers and population heterogeneity: specification and estimation of nonparametric, mixed Poisson model. Criminology 31(4):581–606

Nagin D, Paternoster R (2000) Population heterogeneity and state dependence: state of the evidence and directions for future research. J Quant Criminol 16(2):117–144

Nagin D, Paternoster R (1991) On the relationship of past and future participation in delinquency. Criminology 29(2):163–190

O'Brien RM (1999) Measuring the convergence/divergence of "serious crime" arrest rates for males and females: 1960–1995. J Quant Criminol 15(1):97–114

Ousey GC, Wilcox P, Brummel S (2008) Déjà vu all over again: investigating temporal continuity of adolescent victimization. J Quant Criminol 24(3):307–335

Paternoster R, Brame R (1997) Multiple routes to delinquence? A test of developmental and general theories of crime. Criminology 35(1):49–84

Paternoster R, Bushway S, Brame R, Apel R (2003) The effect of teenage employment on delinquency and problem behaviors. Soc Forces 82(1):297–335

Paternoster R, Dean CW, Piquero A, Mazerolle P, Brame R (1997) Generality, continuity, and change in offending. J Quant Criminol 13(3):231–265

Petras H, Masyn K (2009) General growth mixture analysis with antecedents and consequences of change. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York

Phillips JA, Greenberg DF (2008) A comparison of methods for analyzing criminological panel data. J Quant Criminol 24(1):51–72

Pridemore WA, Chamlin MB, Cochran JK (2007) An interrupted time-series analysis of Durkheim's social deregulation thesis: the case of the Russian Federation. Justice Q 24(2):271–290

Pridemore WA, Chamlin MB, Trahan A (2008) A test of competing hypotheses about homicide following terrorist attacks: an interrupted time series analysis of September 11 and Oklahoma City. J Quant Criminol 24(4): 381–396

Raphael S, Winter-Ebmer R (2001) Identifying the effect of unemployment on crime. J Law Econ 44:259–283

Rosenfeld R, Fornango R, Rengifo AF (2007) The impact of order-maintenance policing on New York City homicide and robbery rates: 1988–2001. Criminology 45(2):355–383

Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27(1):27–55

Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. Justice Quarterly 12(4):625–648

Simpson SS, Bouffard LA, Garner J, Hickman L (2006) The influence of legal reform on the probability of arrest in domestic violence cases. Justice Q 23(3):297–316

Sims CA (1972) Money, income, and causality. Am Econ Rev 62(4):540–552

Smith MD, Devine JA, Sheley JF (1992) Crime and unemployment: effects across age and race categories. Sociol Perspect 35(4):551–572

Stata Press (2005) Stata times-series reference manual release 9. Stata Press, College Station, TX

Stolzenberg L, D'Alessio SJ (1994) Sentencing and unwarranted disparity: an empirical assessment of the long-term impact of sentencing guidelines in Minnesota. Criminology 32(2):301–310

Tracy P, Wolfgang ME, Figlio RM (1990) Delinquency careers in two birth cohorts. Plenum, New York

Weisburd D, Bushway S, Lum C, Yang S (2004) Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42(2):283–321

Witt R, Witte A (2000) Crime, prison, and female labor supply. J Quant Criminol 16(1):69–85

Wu D-M (1973) Alternative tests of independence between stochastic regressors and disturbances. Econometrica 41:733–750