3 The measure of vice and sin: a review of the uses, limitations and implications of crime data
Alexander Tabarrok, Paul Heaton and Eric Helland[1]

INTRODUCTION

In 1670, William Petty, founder of empirical economics, called for the collection of official statistics on 'the number of people, the quality of inebriating liquors spent, the number of unmaryed persons between 15 and 55 years old, the number of Corporall sufferings and persons imprisoned for Crime' all so that that we might know 'the measure of Vice and Sin in the nation'.[2] Petty's call has been answered.

In this chapter we present the major sources of data on crime, brief descriptions and highlights that show the types of data that are available in each source, how these data can be accessed, how they have been in the past and may be used in the future, and some of their limitations. Our goal is to provide a guide to empirical researchers who wish to wade into the waters of vice and sin.

Our primary focus will be on US crime data but we shall also discuss some of the ways in which data across countries can be usefully compared.

In addition to introducing the reader to various data sets, we shall outline methodological considerations of general applicability. Crime data are collected by two methods: reporting and surveys. Each type presents researchers with benefits and drawbacks. One problem with police reports, for example, is that a large fraction of crime is unreported and the data that are reported may frequently be incomplete. Surveys capture some unreported crime but, surprisingly, fail to capture all reported crime and are subject to biases as people fail to remember some events while at the same time incorrectly 'remembering' others. Moreover, answers to surveys can change dramatically with seemingly minor changes in the questions being asked, and different people may understand the same question differently. These and other methodological considerations will be discussed throughout the chapter.

CRIME DATA IN THE USA

We shall cover the following US data sets on crime and related factors: (1) Uniform Crime Reports (UCR), including the Supplementary Homicide Reports, Supplementary Property Reports, and Law Enforcement Employees and Law Enforcement Officers Killed or Assaulted (LEOKA); (2) National Incident-Based Reporting Systems (NIBRS); (3) The National Crime Victimization Survey (NCVS); (4) The Recidivism of Prisoners Released in 1994; (5) Law Enforcement Management and Administrative Statistics (LEMAS); (6) International Crime Victimization Survey; and (7) Geographic Crime Data: Crime Mapping and Spatial Analysis.

The Uniform Crime Reports

Prior to 1930, police reports of crime were neither systematically collected nor compiled, and the definitions of crimes such as burglary, robbery, larceny and so forth were not uniform. The International Association of Chiefs of Police (IACP) created the Uniform Crime Reports (UCR) to collect, compile and standardize data on police reports from agencies across the USA. The UCR is managed by the FBI and became operational in 1930.

The atom of the UCR is the police agency, typically the local police department – other reports such as county, state and national reports are aggregated from the agency data. Agency participation in the UCR is voluntary but participation today is nearly universal with over 17 000 agencies covering 97 percent of the US population reporting to the UCR. Thus the UCR is the most comprehensive database of crime reports in the USA.

UCR are divided into Part I and Part II offenses. The seven original Part I offenses are the violent crimes – criminal homicide, forcible rape, robbery and aggravated assault – and the property crimes – burglary, larceny and motor vehicle theft. The Part I offenses are the most serious offenses and those most likely to be reported to police. Together these offenses are also known as index crimes as they are compiled into the FBI's violent, property and total crime indexes. In 1979 arson was added to the Part I crimes, at congressional request, but the FBI was always concerned about the quality of the arson data and arson is not part of the FBI's crime indexes. In addition, arson is not reported in the most complete and easily available electronic version of the UCR, that produced by the National Archive of Criminal Justice Data (NACJD). We discuss sources of the UCR data further below.

For the Part I offenses, the UCR includes the number of offenses reported, arrests by age, sex, and race of person arrested and clearance rates – an offense is said to be cleared when an arrest for that offense is made and the arrestee is charged and turned over to the court for prosecution. Counts of offenses, arrests and clearances are reported by month for most agencies, although some agencies report only annual totals. In addition, there are two important supplementary reports, the Supplementary Homicide Reports and Supplementary Property Reports, which include additional data and are discussed at greater length below.

Part II offenses are less serious and, to a greater degree than Part I offenses, are often not reported to police. Thus for Part II offenses the UCR only collects data on arrests and the age, sex and race of the person arrested. It is not mandatory for members of the UCR to submit Part II offenses, and thus these data are thus less complete than Part I data. Table.3.1 summarizes.

When aggregated to the state or national level, the basic UCR Part I data on violent and property crime reports are one of the primary sources of data for understanding overall crime trends in the USA (the other primary source being the National Crime Victimization Survey, to be discussed below). Figure 3.1, for example, shows the violent and property crime indexes from 1960 to 2005 based on UCR data. Between 1960 and 1980, crime rates in the USA doubled. For about ten years, crime rates fluctuated before beginning to fall quite dramatically in the early 1990s. Understanding these trends is a key question for economists interested in crime and criminologists (e.g. Levitt, 2004; Blumstein and Wallman, 2000).

Table 3.1   UCR classification of offenses

Part I	Part II
Violent*	Simple assault
Criminal homicide	Curfew offenses and loitering
Forcible rape	Embezzlement
Robbery	Forgery and counterfeiting
Aggravated assault	Disorderly conduct
Property**	Driving under the influence
Burglary	Drug offenses
Larceny	Fraud
Motor vehicle theft	Gambling
Arson (as of 1979 and not usually indexed)	Liquor offenses
	Offenses against the family
	Prostitution
	Public drunkenness
	Runaways
	Sex offenses
	Stolen property
	Vandalism
	Vagrancy
	Weapons offenses
Available data: Offenses reported, arrests by age, sex, and race of person arrested and clearance rates.	**Available data**: Arrests by age, sex, and race of person arrested.

Notes:
*   **Supplementary Homicide Reports** add age, sex and race of victims, and (where known) offenders, victim–offender relationships, weapons used and circumstances surrounding the homicides.
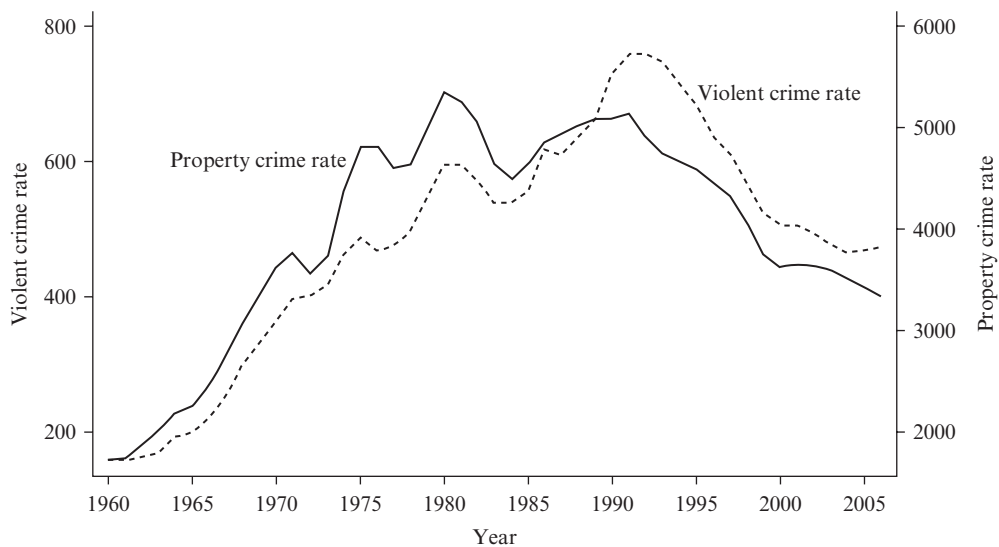** **Supplementary Property Reports** include information on the value and type of property stolen during a murder, rape, robbery, burglary, motor vehicle theft or larceny.

The UCR data have been used in hundreds of studies of crime in the USA, especially when broken down by state and county, which allows for better testing of hypotheses (see this volume for many citations).

The Supplementary Homicide Reports (SHR) portion of the UCR includes information on the race, age and gender of the offender (where known) and the victim. In addition, there are data on the relationship between the offender and victim (stranger, boyfriend, husband etc.) and the circumstances of the homicide (robbery, lovers' triangle, brawl etc.).[3]
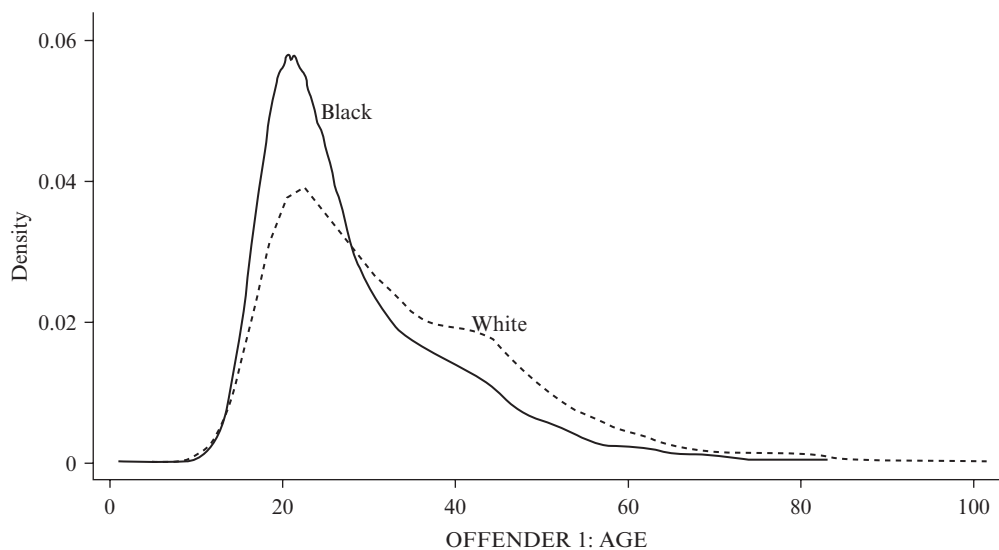
Figure 3.2, for example, uses data in the SHR to estimate and graph the density function for the age of homicide offenders by race.[4] The peak probability of offending for both blacks and whites is around age 20, but black homicide offenders are clustered around age 20 to a greater extent than white offenders – indicating, as is well known, the very high rate of offending among young black men. Victim ages tend to follow offender ages, so Figure 3.2 also suggests high rates of victims among young blacks.

Figure 3.3, also based on data from the SHR, plots the percentage of white and black young (aged 14–24) males among the population and among homicide offenders from
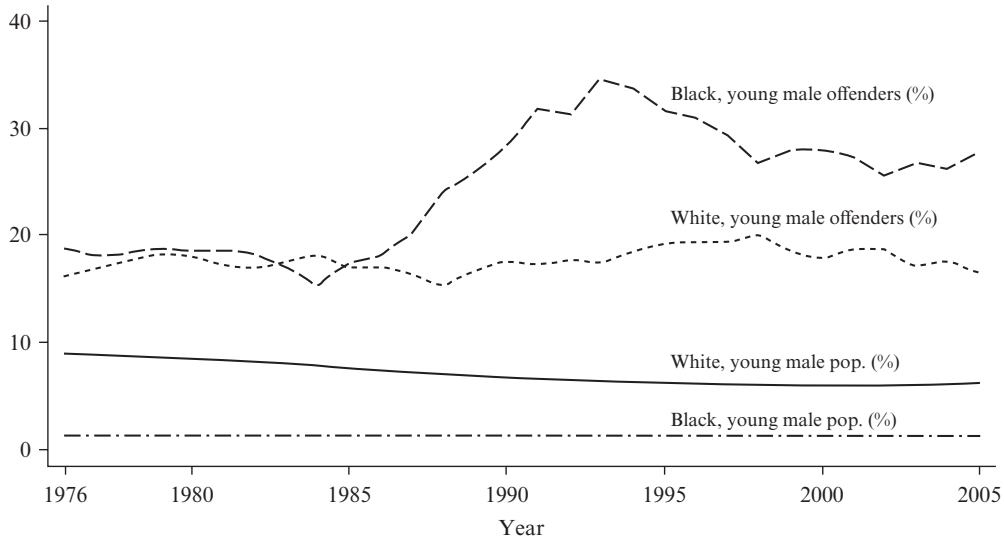
Figure 3.1   Violent and property crime rates per 100 000 population, 1960–2006

Note: For incidents with multiple offenders the first offender's age was used.

Figure 3.2   Distribution of homoicide offender ages for black and white offenders, 2004

Source: Bureau of Justice Statistics, *Homicide Trends in the United States, 2007*. Available online at www. ojp.usdoj.gov/bjs/homicide/homtrnd.htm#contents.

Figure 3.3   White and black males aged 14–24 as percentage of population of homicide offenders, 1976–2005

1976 to 2005. In 1989, for example, young black males committed nearly 26 percent of all homicides while accounting for just 1.2 percent of the population. In the same year, young white males committed around 17 percent of all homicides while accounting for 7 percent of the population. Thus both young black and white males committed homicides in greater proportion than expected by population alone, but the ratio was far higher among young, black males. Over the 1976–2005 period the ratio of offender percentage to population percentage for young black males averaged 20.7 and for young white males 2.5. Also evident in the figure is the large increase in black offending rates from 1985 to 1995 during the so-called crack epidemic (Fryer et al., 2005).

Supplementary Property Reports (SPRs) include information on property stolen during a murder, rape, robbery, burglary, motor vehicle theft or larceny.[5] The SPRs contain data on the basic nature of the crime (e.g. robbery is subdivided into highway, common house, gas station, chain store, residence, bank or miscellaneous robbery), the monetary value of the stolen property, the type of property stolen (e.g. currency, jewelry, firearms, livestock etc.). For motor vehicles there is also information on the amount of property recovered.

In addition to crime data, the UCR collects basic information on law enforcement officers, including the number per agency and gender, and information on law enforcement officers killed or assaulted.

Sources of UCR data
The UCR collects an immense amount of data and there is no single source from which all the data can easily be extracted. The first place to look is the Bureau of Justice

Statistics (BJS) online data page, http://www.ojp.usdoj.gov/bjs/dtd.htm. The BJS offers quick and easy access to Part I offense numbers and rates by agency (with population greater than 10 000) since 1985. Aggregated data by state or for the nation as a whole are available for the years 1960–2006 continuing. A variety of other information on offenses, arrests and demographics by various aggregations, including counties, can be found on the BJS website.

Another source for quick access to various tabulations of the UCR data, including aggregations to county, state and region, is the FBI's annual report *Crime in the United States* (1995–2007, continuing). A longer time series of UCR data (along with NCVS and NIBRS data described below) can also be obtained in a fairly user-friendly format through the Data Cubes application developed by the National Consortium on Violence Research (NCOVR, www.ncovr.heinz.cmu.edu). Simple online data analysis is also available for some series at the Office of Juvenile Justice and Delinquency Prevention. For more detailed analysis, the National Archive of Criminal Justice Data (NACJD), a part of the Interuniversity Consortium for Political and Social Research (ICPSR)*,* contains the most complete yet still relatively accessible subset of the UCR data in electronic form.

At the NACJD, the series *Uniform Crime Reporting Program Data [United States]: Offenses Known and Clearances by Arrest* includes agency data on offenses reported, arrests and clearances for Part I offenses for the years 1966–2004 (continuing). In this series the data are broken down according to whether the arrestee is under or over 18, but it does not otherwise include the age, sex or the race of the arrestee. Moreover, this series does not include information on Part II offenses.

Supplementary Homicide Reports are available in the series *Uniform Crime Reporting Program Data [United States]: Supplementary Homicide Reports* for the years 1975–2004 (continuing). Supplementary Property Reports are available in the series *Uniform Crime Reporting Program Data [United States]: Property Stolen and Recovered* for the years 1966–2004 (continuing).

Information on law enforcement officers, along with detailed information on law enforcement officers killed or assaulted, is available in *Uniform Crime Reporting Program Data [United States]: Police Employee (LEOKA) Data* for the years 1975–2004 (continuing).

The NAJCD also aggregates the agency-level data to the county level. County-level data are very useful if one wants to examine correlates of crime because a wide variety of demographic data is available from the Census by county. The county data include information on Part I and Part II offenses and also arrests by those under and over 18, but does not otherwise include age, sex or race characteristics of offenders. We shall discuss problems with UCR data further below, but it is important to note at this point that not all agencies report every month, so the NAJCD must impute missing data to aggregate to the county level. The imputation algorithm changed in 1994 and thus comparisons pre and post-1994 have to be made with care. County-level data are available from 1977 to 2005 (continuing).

All of these data can be accessed conveniently from the NAJCD web page, http://www.icpsr.umich.edu/NACJD/ucr.html.

Data on the age, sex and race of arrestees for both Part One and Part Two crimes may be obtained in the series *Uniform Crime Reporting Program Data [United States]:*

Arrests by Age, Sex, and Race, but this series exists in electronic form only for 1994–2006 (continuing). The data are aggregated by month. It is possible to create cross-tabulations by age and sex, but not by age and sex with race. That is, one can find how many arrests there were in a particular month of females aged 25–29 or how many blacks were arrested but not how many arrests there were of black females let alone black females aged 25–29 (it is possible to distinguish between juvenile and adult arrests by race, however).

We have focused on sources of electronic data because of the importance of ease of access. It should be noted, however, that the FBI has more data than are available electronically and in our experience they will respond to requests for additional data if they can. One of the authors, for example, casually asked for additional FBI data and was surprised to find six large boxes of printouts delivered to his office several weeks later.

Problems and issues with UCR data: conceptual and practical

The primary conceptual issue with the UCR data is the hierarchy rule, which requires that when a single incident involves multiple Part I offenses, the agency report that crime, and only that crime,[6] which is highest on the list of Part I offenses (called the hierarchy list). The FBI's UCR *Handbook* provides a useful example of the hierarchy rule:

> **Situation**: A burglar broke into a home, stole several items, and placed them in a car belonging to the owner of the home. The homeowner returned and surprised the thief, who in turn knocked the owner unconscious by hitting him over the head with a chair. The burglar drove away in the homeowner's car.

> **Applying the Hierarchy Rule:** A Burglary – Forcible Entry (5a), Larceny-theft (6), Robbery – Other Dangerous Weapon (3c), Aggravated Assault – Other Dangerous Weapon (4d), and Motor Vehicle Theft – Auto (7a) occurred in this situation. After classifying the offenses, the reporting agency must score only one offense – Robbery – Other Dangerous Weapon (3c) – the crime appearing first in the list of Part I offenses.

The hierarchy rule is not as constraining as it first appears because it applies only to single incidents, not to multiple offenses separated in time or place but committed by the same criminals. For example (again from the UCR *Handbook*), a gunman surprises a man and a woman in a parked car. He kills the man and abducts the woman driving her across town to a secluded spot where she is raped. The murder and the rape are separated in time and space and are considered as two incidents – thus the hierarchy rule would not apply and both crimes would be reported.

The hierarchy rule was used to simplify the collection and storage of crime data before computer records became common. With computer records, it has become possible to eliminate the hierarchy rule and collect more data about each crime incident. The National Incident-Based Reporting System begins this task and is described further below.

The primary practical problem with UCR data is that a surprising amount of them are missing or inaccurate. The national, state and county data are aggregated using data from approximately 17 000–18 000 reporting agencies (typically police departments). As noted earlier, agency data are the atom from which all the other data are built. Appropriately enough, therefore, an uncertainty principle applies to UCR data: the closer one gets to the atom, the greater the uncertainty.

Table 3.2   States and years in which supplementary Homicide Reports are not available or severely undercounted, 1980–2005

State	Data years not available
Alabama	1999
Delaware	1994, 1995
District of Columbia	1996, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005
Florida	1988, 1989, 1990, 1991, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005
Illinois	1984, 1985, 1987
Iowa	1991
Kansas	1988, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000
Kentucky	1987, 1988, 1994, 1999, 2000, 2001, 2002, 2003
Louisiana	1991
Maine	1991, 1992, 1993
Montana	1982, 1986, 1987, 1990, 1993, 1994, 1996, 1997, 1998, 1999, 2001, 2002
Nebraska	1987, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2005
New Hampshire	1997
New Mexico	1981
North Dakota	1991
South Dakota	1997
Vermont	1981, 1982, 1983
Wisconsin	1998

Note: Severely undercounted is defined as collected at a rate such that total homicides are more than 50% below the rate predicted by the FBI.

Source: Puzzanchera and Kang (2008).

At the national level, random errors are likely to cancel and only a small fraction of the total data needs to be imputed, so the data are likely to accurately reflect national trends. As we disaggregate to the state level, then missing data become important. Table 3.2, for example, shows the states and years between 1980 and 2005 for which Supplementary Homicide Reports are not available or are so low as to require imputation. Note that it is not just SHRs that may be unavailable – this is just an illustration of the types of problems that exist in the UCR data.

If we disaggregate further to the county level, then more data are missing and the NAJCD imputes as necessary. It is important to note that data are not missing at random – smaller counties, for example, are likely to have more missing data than larger countries. The current imputation procedure is as follows: If an agency reports for two or fewer months, crime for that jurisdiction is estimated based on crime reported by agencies within the same state[7] with the same population size. This estimation procedure ignores differences in income levels and distribution, population density, ethnic and racial composition, and other important factors. The rape rate in urban and suburban jurisdictions is more than 20 percent higher than in rural areas; this, however, would not be accounted for under the current imputation methods.

If an agency reports for more than two but fewer than twelve months in a particular

year, the number of incidents reported is divided by the number of reporting months and multiplied by 12. The methodological soundness of such simple imputation[8] is questionable; for example, rapes are committed at a disproportionately high rate during the late spring and summer months. In addition, this imputation method was implemented in 1994. A similar but simpler method was used previously to 1994, and as a result, the data collected before 1994 are not necessarily comparable to data collected since. It should be noted that county-level data are provided solely for research purposes and are not official UCR releases.

Regarding the county-level offense data, Maltz and Targonski (2002, 2003) argue 'that in their current condition, county-level UCR crime statistics cannot be used for evaluating the effects of changes in policy'. Lott and Whitley (2003) offer a more optimistic argument but recognize that account must be taken of the sometimes poor quality of the county data.[9]

As we disaggregate further, getting closer to the atom, we find even more missing data – Maltz (1999), for example, found that between 1992 and 1994 only 64 percent of agencies filed reports for each of the 36 months and 19 percent filed no reports at all during this period. In addition, the agency data contain other peculiarities such as spikes. Spikes might represent crime waves but may also occur when the person responsible for submitting UCRs becomes overworked and in response to a backlog submits August and September reports as September reports, creating a September spike.

When using the agency-level data, especially for smaller agencies, we must bear in mind the wisdom of the British economist Sir Josiah Stamp (1929, p. 258), who wrote:

> The government are very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn pleases.

In short, it is highly recommended that a researcher using the county or agency should consult the previously cited papers by Maltz (1999), Maltz and Targonski (2002, 2003) and Lott and Whitley (2003) for guidance on the data quality and some techniques for minimizing inference problems. In addition, Maltz has recently produced an Excel file of monthly UCR data from 1960 to 2004, for over 17 000 police departments and 26 crime categories or subcategories. The data have been cleaned to some extent but their primary use is not for further analysis (extracting the data from the Excel file can be cumbersome); rather, Maltz has provided an easily accessible visual guide to what data are missing or otherwise questionable. Researchers should consult this data set to get a feel for the types of issues they will need to confront in more refined analysis. The Maltz data are available at http://sociology.osu.edu/mdm/UCR1960-2004.zip.

National Incident-Based Reporting Systems (NIBRS)

Development of the National Incident-Based Reporting System (NIBRS) began in 1984 due in large part to growing dissatisfaction with the narrow scope of information detailed in the UCR. The system has been operational since 1988, but has been implemented only in a limited number of jurisdictions. Table 3.3 reports the total number of crime incidents and the number of agencies participating in NIBRS over time. Participation in the

Table 3.3   NIBRS data availability by year

Year	Group A incidents	Group B incidents	Participating agencies
1995	837 014	318 524	1255
1996	1 064 763	387 622	1487
1997	1 426 978	531 097	1738
1998	1 822 675	713 057	2249
1999	2 157 326	830 071	2852
2000	2 841 523	1 006 424	3365
2001	3 232 281	1 044 178	3611
2002	3 455 589	1 126 216	3809
2003	3 637 432	1 154 498	4287
2004	4 083 571	1 296 557	4525
2005	4 614 054	1 457 435	4682
2006	4 906 781	1 540 038	4841

NIBRS program is far lower than UCR participation not only because it is relatively new, but also largely because of the comprehensive nature of the program. The benefits of more detailed record-keeping may predominantly accrue to researchers and scholars, while the costs are primarily borne by taxpayers and law enforcement agencies. The survey is not a representative sample of the USA; participation has been greatest among small and medium-sized law enforcement agencies. Given that compliance is voluntary and relatively low, researchers should use caution in drawing inferences using NIBRS data, especially those that pertain to absolute measures of crime.

Although less geographically comprehensive, NIBRS provides several improvements over the UCR. Whereas the UCR only reports aggregate counts of crimes by agency, NIBRS takes an individual crime incident as the basic unit of observation, providing a much finer-grained portrait of crime in a particular jurisdiction. NIBRS differentiates between attempted and completed crimes, and permits linkages between a single incident and multiple individuals and offenses, eliminating the need for a hierarchy rule. Crimes are classified according to 22 offense categories encompassing 46 discrete (Group A) offenses; for these offenses all reported crimes are included in the data. Additionally, there are 11 Group B offenses for which only arrest information is reported.[10] Each incident is linked to information about the reporting agency, particular offenses involved and their circumstances, victims/offenders, property, and arrestees if an arrest is made. To preserve confidentiality, NIBRS includes a fairly limited set of demographic identifiers for individuals (age, gender, race), and particular individuals are not linked across crime incidents.

NIBRS permits researchers to focus on more narrowly defined crime types for purposes of analysis. For example, NIBRS allows identification of crimes occurring on particular dates or days of the week; crimes involving the use of weapons, alcohol, or drugs; and multiple-offender crimes. Because offense and arrest records are linked, NIBRS facilitates analysis of time to arrest and factors influencing arrest. The compilation of victim data under NIBRS also permits research to examine crimes directed at particular types of individuals, including domestic crimes.

NIBRS data are available from the Inter-University Consortium for Political and

Social Research (ICPSR). The data collected by the FBI are recorded across multiple data files, posing significant difficulties for researchers who might lack adequate computing resources to deal with such large data sets. Moreover, it requires the user to be well versed in manipulating the complex file types. Data files may be downloaded for use in SAS, SPSS and Stata formats through the ICPSR at http://www.icpsr.umich.edu/NACJD/NIBRS/. Full-level files are available for the years 1995–2006. Extract files, which simplify analysis by organizing data along four dimensions: incident, victim, offender and arrestee, are available for the years 1997–2005.

National Crime Victimization Survey (NCVS)

Formerly known as the National Crime Survey, the National Crime Victimization Survey (NCVS) has been in operation since 1973. The primary virtue of a victimization survey is the possibility of capturing crimes that are not reported to the police and providing more detail about victims and circumstances of the crime.[11] In the 2006 survey, victims say that they report to the police just less than half (48.8 percent) of all violent crime and only 37.8 percent of property crime (US Department of Justice, 2008). Thus differences between the NCVS and the UCR can be large, although they have diminished over time (McDowall and Loftin, 2007). The primary data on who is victimized by crime by age, race, gender, income, household size and so forth come from the NCVS. So if you want to compare the victimization rate for violent crime for white males aged 25–34 with the victimization rate for black males of the same age, the NCVS and the extensive tables produced by the Bureau of Justice Statistics is the place to go (the rate for white males is 34.2 and for black males 60.8 per 1000 persons in each group).

The US Census Bureau administers the survey on behalf of the Bureau of Justice Statistics. The NCVS employs a complex, stratified, multi-stage cluster. Household selection occurs by a rotated panel design where each household is interviewed seven times at six-month intervals. Respondent households are designed to be a nationally representative sample of US residences.

Data can be downloaded for use with SAS, SPSS, and Stata[12] from ICPSR at http://www.icpsr.umich.edu/NACJD/NCVS/. Incident-level concatenated files and full files are available for the period 1992–2005. In order to facilitate analysis, record-type files sort annual data into four classifications: address, household, person and incident; they are available for the period 1992–2006. Data on metropolitan statistical area (MSA) are available for the period 1979–2004. MSAs are composed of the core counties for the 40 largest metropolitan areas in the country over the period[13]. This extract contains two files: a weighted, person-based file and a weighted, incident-based file. In addition, unbounded survey data are available for 1999–2004; a school crime supplement is available for 1995, 1999, 2001, 2003 and 2005.

When it began in 1972, the NCVS sampled 72 000 households, but over time pressures to reduce costs have meant reductions in the number of questions, the types of people questioned (in particular a commercial sample was cut in 1977), and the sample size, which today is just under 40 000 households (Groves and Cork, 2008). Data are collected by means of both face-to-face interviewing (FTF) and computer-assisted telephone interviewing (CATI), although CATI is becoming more common, once again because of tight budgets.

One major obstacle to obtaining reliable results from victimization studies is the cognitive bias known as telescoping, reporting events that occurred prior to the time period in question as having occurred during that time period. Victimization rates can easily be doubled, for example, if respondents report any incident that occurred in the last year as occurring in the last six months. To control for telescoping, NCVS tries to interview respondents every six months over the course of 3.5 years for seven interviews in total. The results of the first interview are not counted in victimization statistics (until recently, see below) but are used to bound subsequent interviews – thus if the same incident is described in a subsequent interview, the interviewer can ask the respondent to clarify whether this incident is indeed a new incident.

Unfortunately, the panel design uses physical residences, not unique families, as the units to which bounding is applied. Thus, if a family in the panel moves, that family does not remain in the panel, but the new occupants of the residence are asked to join it. Moreover, the crimes reported in the new family's first interview are counted among official survey results because this is a subsequent interview for the residence! Since families move quite often, a significant fraction of interviews are in fact unbounded. Biderman and Cantor (1984) found that 18 percent of all interviews used in the compilation of official statistics were in fact unbounded. The inclusion of unbounded interviews *per se* tends to increase victimization statistics. On the other hand, if movers experience greater crime rates than non-movers, the failure to track families could reduce estimated victimization rates. In 2007, due to budget cuts it was decided to include data from the first interview for all families, so bounding is likely to become a more serious problem in the future (Groves and Cork, 2008).

In an attempt to minimize conceptual ambiguities, the NCVS underwent a significant redesign in 1992 as screening questions were expanded and clarified. The redesign has compromised the continuity of the survey to such an extent as to render the pre- and post-redesign data almost useless for purposes of comparison. Cantor and Lynch (2000) found that the changes substantively altered the responses given, leading to an increase in reported crime between 50 and 200 percent, depending on the crime.

Since the redesign, five of the seven interviews have been conducted using the CATI method. The CATI method has been shown to yield reported crime incidents 15–20 percent higher than FTF interviews (Mosher et al., 2002). The authors attribute this effect to greater standardization and monitoring associated with the CATI process; however, it might be that respondents feel more free to answer candidly on the phone. Other research, however, has found little difference between the responses offered during CATI rather than FTF interviews (van Dijk and Mayhew, 1992; Lynch, 2006; Catalano, 2007).

Moreover, survey definitions may be more or less inclusive of various acts as compared to legal definitions. Biderman and Lynch (1991) and O'Brien (1985) indicate that survey definitions of crime tend to be broader than legal definitions – much of that which is counted as crime in surveys is not crime by legal definition.

While many crimes are not reported to law enforcement agencies, survey data may not necessarily be a less biased estimate of total crime than police reporting. Persons with a college degree report more assaults than those with only an elementary education, even though one would expect assaults among the latter to exceed those among the former (Mosher et al., 2002). It is not readily apparent whether this should be interpreted as

college graduates actually experiencing significantly higher rates of victimization, or as tending to be more willing than others of lower educational attainment to classify various activities as constituting criminal activity.

There is also evidence that, in the face of persistent questioning by interviewers, respondents will detail the crime experiences of friends, family and neighbors. On the other hand, Turner (1972) found that only 63 percent of all cases of robbery, assault and rape in police records were reported on victimization surveys; 76 percent when the offender is a stranger to the victim and only 57 percent of cases where the victim knows the offender.[14]

Groves and Cork (2008) offer a useful history of the NCVS, including a summary of survey problems and issues.

Recidivism of Prisoners Released in 1994

One major limitation of the data sources discussed thus far is that they do not allow researchers to track an individual's response to policy changes. Ideally researchers could utilize a panel of individuals and estimate the impact of more stringent or longer periods of incarceration, or the impact of employment opportunities on the likelihood that an individual commits a crime. Such a panel would allow researchers to control for a number of unobservable factors that potentially impact propensity toward criminal activity. The most common panel data set used for such studies is the National Longitudinal Survey of Youth, which contains questions on criminal activity (see, for e.g., Freeman, 1996; Grogger, 1998). One problem with household surveys, however, is that criminal activity, especially serious criminal activity, is a relatively rare event, so few individuals in any household survey will have had extensive interactions with the criminal justice system.

One solution to this problem is to examine individuals who have already been incarcerated. Witte (1980) provides an early example of this approach in a study using individual data on men released in North Carolina. The difficulty for researchers interested in this approach is that the data must often be obtained by the researcher directly from the criminal justice system, and typically this fact necessitates that the study examine only one jurisdiction and a relatively small sample (also see Tauchen, Chapter 2 in this volume).

The exception to this rule is a data source assembled by the Bureau of Justice Statistics (BJS) called the Recidivism of Prisoners Released in 1994 (ICSPR 3355). In 1997, the BJS collected data from 15 departments of corrections and compiled the complete criminal history of a sample of prisoners released in 1994. The data on each released prisoner come from the officially recorded history captured in the FBI and State RAP (Records of Arrests and Prosecutions) sheets.

Specifically, the BJS assembled a representative sample of 38 624 prisoners drawn from the 302 309 records supplied by the 15 state corrections agencies. The data were then cross-checked with similar information obtained from the FBI. Thus constructed, the data contain information from the releasees' first arrest to the end of the sample period in 1997.[15] Table 3.4 shows the number of releasees for each state in the data.

The file is arranged by released prisoner and contains up to 99 'arrest cycles'. Each cycle is determined by the date of arrest and includes details on the crimes committed (i.e. arrested for assault and burglary if the arrest occurred at the same time) and the

Table 3.4   Tracked prisoners by state

State	Released prisoners
Arizona	2000
California	7183
Delaware	721
Florida	2893
Illinois	2615
Maryland	2117
Michigan	2315
Minnesota	1929
New Jersey	2289
New York	2639
North Carolina	2314
Ohio	2664
Oregon	2292
Texas	2550
Virginia	2103
Total	38624

Source: Bureau of Justice Statistics (1994).

adjudication (i.e. pleaded guilty and given six months). The data on arrest and adjudi-cation are very detailed as each charge of the arrest has multiple codings for the crime committed, whether the charge was a felony or a misdemeanor, as well as information on the victims.

Arrest cycles begin with the prisoner's first arrest and track through the end of the three-year observation period. Thus a prisoner who was first arrested in 1984 might have multiple arrests, convictions and incarcerations prior to the incarceration that resulted in the 1994 release. After release he may again have multiple arrests and adjunctions. Thus researchers can construct a released prisoner's criminal history prior to his 1994 release as well as tracking him over the three years subsequent to release. Because criminals are very likely to recidivate, the data can be used to construct a panel of individuals who have had an extensive interaction with the criminal justice system and are likely to be rearrested during the sample period.

The data are not a complete record of all offenses. Most obviously they do not include crimes committed by the releasee for which he is not arrested. Moreover, they do not include juvenile records or petty offenses that do not, typically, appear on the RAP sheet.

There are several larger methodological issues with the data. First, they contain little demographic information beyond race, age and gender. Most importantly for studies of deterrence they contain nothing on education, employment or earnings. Moreover, while the data on criminal history and post-release criminal activity are extensive and rarely missing, the information that is missing typically occurs earlier in the releasee's criminal history. There are also tantalizing clues about the releasee's experience in prison, for example, there are entries for whether the individual is a drug or alcohol abuser and whether the individual had drug or alcohol treatment courses or took vocational courses

while in prison. But unfortunately these data are missing for most states. This data set, however, has not been used very often so the enterprising researcher may be able to exploit the information that does exist.
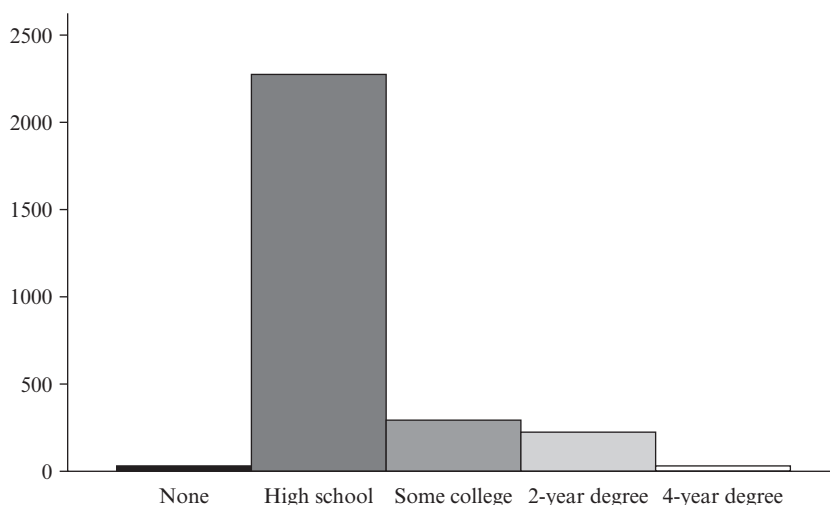
Helland and Tabarrok (2007) use the data to evaluate California's three-strikes law. Using the data set's information on prior criminal history, they are able to construct how many strikes a criminal had when released in CA in 1994, just as the three-strikes law was put in place (previous strikes count as strikes under CA's law). To test the effect of the three-strikes law, Helland and Tabarrok compare the post-sentencing criminal activity of criminals who were convicted of two strikeable offenses (and thus were eligible for a third strike) with those who were tried for a second strikeable offense but convicted of a non-strikeable offense (and thus were not eligible for the third strike). Thus the randomization of trial outcomes is used as a quasi-experiment to create comparable criminals. Helland and Tabarrok (2007) find that the three-strikes law reduced the criminal propensity of eligible third strikers by 17–20 percent compared with similar individuals not subject to the law. Using a similar method, Tabarrok and Helland (2009) find that the three-strikes law did not induce significant movement of criminals from CA to other states; thus spillovers were minimal.

Lemas: Law Enforcement Management and Administrative Statistics

The Law Enforcement Management and Administrative Statistics (LEMAS) program collects data about the management and administration of state and local law enforcement agencies throughout the country. Managed by the US Department of Justice, Bureau of Justice Statistics, the first program survey was conducted in 1987, with subsequent installments in 1990, 1993, 1997, 1999, 2000 and 2003.[16] Each year, local agencies with more than 100 officers and state police departments are automatically surveyed, with a nationally representative sample of smaller agencies also surveyed. Data are formatted for use with SAS and SPSS; the 2003 data are also formatted for use with Stata. Data for each installment can be downloaded from ICPSR at http://www.icpsr.umich.edu/cocoon/NACJD/SERIES/00092.xml.

The survey is administered to approximately 3000 law enforcement agencies every three to four years. Larger agencies are given a longer questionnaire form with more detailed questions. In 2003, the response rate was 90.6 percent, with 2859 agencies reporting. LEMAS provides in-depth information (618 discrete variables) about law enforcement agencies, including, among other things: the [employment status] of officers, the demographic composition of law enforcement agencies, the facilities, tools and technology available to officers, the modes of transportation used and the extent to which each is employed, and whether they use surveys to gauge public perception. Questions vary across years of the survey – for example only the 1987 survey contains information regarding litigation against police, while more recent surveys include a number of questions about community and problem-based policing that are absent from earlier surveys. LEMAS does not collect data about crime incidents or overall crime levels, although LEMAS data do include FBI agency codes, which allow LEMAS data to be linked to crime data from the UCR or NIBRS.

These data contain a great deal of information that may be of interest to researchers. Garicano and Heaton (2006), for example, use LEMAS data to evaluate the effectiveness

Source: LEMAS (2003).

Figure 3.4   Law enforcement and education: minimum educational requirements in 2859 US agencies

of IT, including innovations such as Compstat in crime fighting. LEMAS also contains a great deal of data on police hiring and characteristics. For example, of the 2859 participating agencies, 27 do not require a criminal background check prior to employment. Additionally, 712 agencies do not require the administration of a psychological examination prior to employment. Requisite education for employment in many agencies is quite low. As shown in Figure 3.4, nearly 80 percent of reporting agencies require no more than a high-school education and 38 agencies have no educational requirement whatsoever.

Although LEMAS is designed as a cross-sectional survey, because individual agencies are identified and larger agencies are automatically sampled, it is also feasible to combine data from multiple survey years and track agencies over time. Figure 3.5, for example, draws from several LEMAS surveys to plot trends for several large metropolitan departments in the number of service calls, one proxy for police workload. Although calls for service were fairly stable in Pittsburgh and Seattle, they grew dramatically in Las Vegas and Houston, two cities that were experiencing substantial population growth over the sample period. Cincinnati also saw large increases in police workload despite losing population during the 1990s. The major limitations to longitudinal analysis of LEMAS data are the changes in the survey questions and probabilistic sampling of smaller agencies.

International Crime Data

The International Crime Victimization Survey (ICVS) became operational in 1989, with subsequent installments in 1992, 1996, 2000 and 2004–5. It is the most ambitious and comprehensive attempt to minimize the myriad of difficulties associated with making cross-country comparisons of crime victimization. Data are available in SPSS
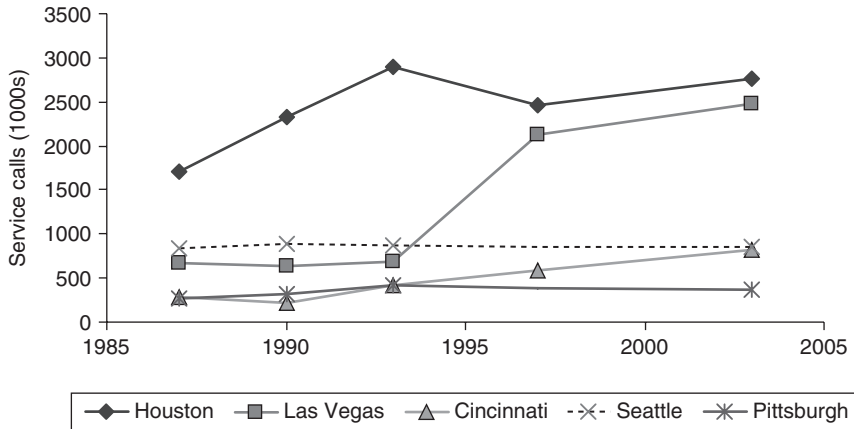
Figure 3.5   Trends in service calls for selected metropolitan police departments

and SPSS portable formats and can be downloaded from http://www.unicri.it/wwd/ analysis/icvs/data.php. Sample questionnaires are available at http://rechten.uvt.nl/ icvs/#Questionnaires.
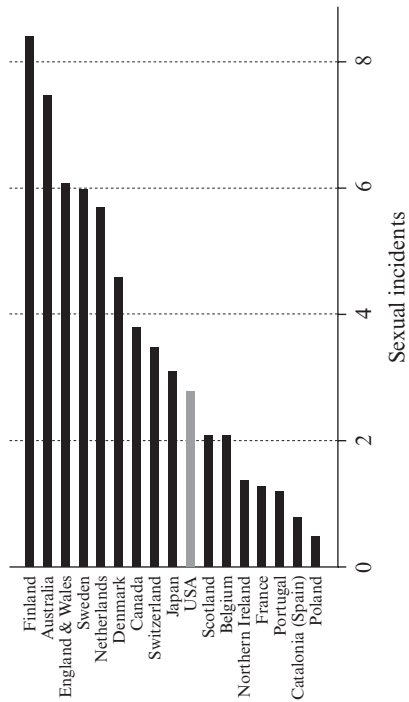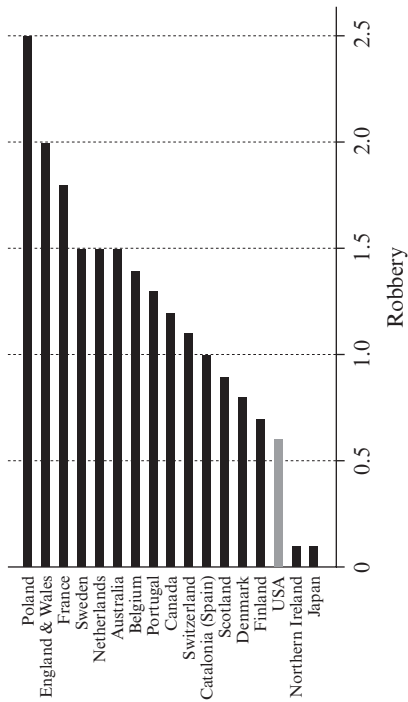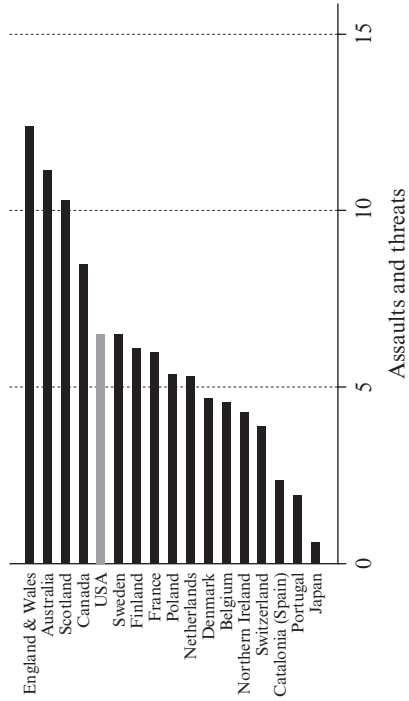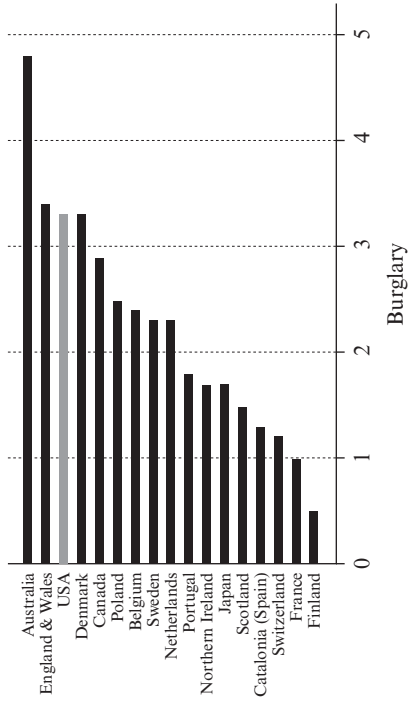
The 1989 sweep included 14 industrialized countries; several city-specific surveys were also administered in the first installment. The ICVS uses a standard sample population size of 2000 individuals for each country. The most recent installment included responses from individuals in 30 countries as well as 33 city-specific surveys. Over the course of the survey, more than 300 000 individuals from 78 countries have participated in the ICVS. Five countries – Canada, England and Wales, Finland, the Netherlands and the USA – have participated in all five rounds of the ICVS. Whether an individual chooses to report a crime to authorities depends on several factors: the severity of the crime, the monetary value (if any) of the criminal act, police responsiveness and perceptions of corruption, social norms, and the opportunity cost of the victim's time. It is likely that the rate at which crime is reported varies significantly both within and between countries. Such considerations would favor the use of well-designed surveys for estimating crime levels at an international level.

Yet cultural differences in sensitivity to violence may impact the results of survey data as well. The inclusion of a growing number of developing nations in each subsequent installment poses a challenge to survey designers; they take great care to maintain cross-country comparability and standardization by ensuring that translations are as precise as possible. Nevertheless, as with all surveys, issues of design, response rate, memory and so forth are important (see Mayhew and van Dijk, 1997 and Alvazzi et al., 2000 for further discussion).

Unlike the NCVS, the ICVS asks respondents about consumer fraud victimizations. The survey also attempts to measure individuals' experiences with government corruption.
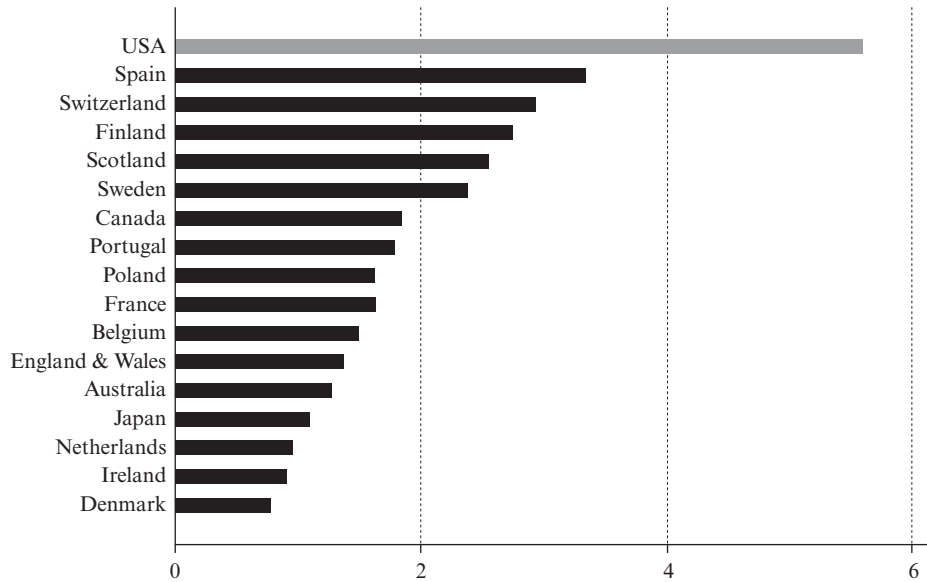
Crime and homicide: comparing the USA to other countries
Figure 3.6 compares robbery, burglary, sexual incidents (assaults), and assaults and threats in the USA and a large number of other industrialized countries using data from

Source: Van Kesteren et al. (2000).

Figure 3.6 Crime in the USA and other industrialized countries (offenses per 100 inhabitants, 2000)

Country	Rate

(chart: horizontal bar chart showing homicide rate per 100,000)

USA, Spain, Switzerland, Finland, Scotland, Sweden, Canada, Portugal, Poland, France, Belgium, England & Wales, Australia, Japan, Netherlands, Ireland, Denmark

x-axis: 0　2　4　6

Note:　Definitions of homicide may vary by country.

Source:　Various by country as compiled by Wikipedia.

Figure 3.7　Homicide rate by country (rate per 100 000, circa 2000–2007)

the ICVS. Perhaps surprisingly, the USA is, in these respects, a country with a low to medium rate of crime.

The ICVS does not compare homicide across countries, and the definition of homicide can vary. Unfortunately, there is no central source of comparable homicide data, but on any reasonable comparison the USA does have a high homicide rate relative to other industrialized nations, as shown in Figure 3.7. Thus it is the US homicide rate not the US crime rate that stands as the outlier.

While homicide in the USA is extremely high as compared to other industrialized countries, Colombia's murder rate is more than 14 times that of the USA, and many less industrialized countries have murder rates more than double that of the USA.

GEOGRAPHIC CRIME DATA: CRIME MAPPING AND SPATIAL ANALYSIS

Crime mapping has a long history, dating at least to the cartographic school of Quetelet (1831) and Mayhew (1861), but crime mapping did not become an important tool in crime fighting until the 1990s, when police departments, led by the NYPD, made the weekly collection and analysis of data the central organizing principle of police strategy. Many large police departments now have statistical units that collate, manage and analyze crime databases. With the rise of the Internet it became a natural and simple

Figure 3.8   Geographic crime patterns

procedure to export crime data to the Web for use by the public. The LA, Chicago, Washington, Oakland and Baltimore police departments, among others, offer online mapping of crime incidents, usually from the past several months. Websites such as www.SpotCrime.com collate data from many police departments around the USA and produce maps on request using online tools such as Google Maps. SpotCrime.com, for example, produced the map in Figure 3.8 showing approximately one month's worth of arson, assault, robbery and shootings in Oakland, CA.

   Maps like those in Figure 3.8 can suggest interesting avenues of research, but for real analysis it is important to integrate geographic data with other types of data. Integrating geographic and 'ordinary' data, however, is not easy. A variety of formats for geographic data exist and no single software package is capable of mapping, manipulating and analyzing all the types of geographic and non-geographic data that a typical researcher will be interested in. A newcomer to this field should be prepared to be frustrated.

Table 3.5   Homicides, Robberies and Demographics in Washington, DC by census block group

	Homicides	Robberies
Constant	1.962	33.854
	(0.523)**	(10.217)**
African Americans	0.001	0.028
	(0.001)*	(0.011)**
Hispanics	−0.001	0.117
	(0.001)	(0.015)**
Per capita income (1000s)	−0.048	−0.046
	(0.01)**	(0.197)
Youths aged 5–17	0.013	−0.007
	(0.002)**	(0.037)
Observations	429	429
R-squared	0.51	0.21

Notes:
Standard errors in parentheses.
* Significant at 5%
** Significant at 1%
Data include all reports from January 1997 to July 2003.

We offer a brief guide to mapping and analysis in Stata and GeoDa, the former an extensive statistical package often used by economists and the latter a free software product for analyzing spatial data.[17]

Crime data from police departments will typically come in the form of addresses (typically redacted to the block level). The first step in analysis is to geocode the addresses, that is assign geographic identifiers to each address. For example, we obtained data on over 285 000 crimes from January of 1997 to July of 2003 from the Metropolitan Police Department of the District of Columbia. For each address, we paid Geocode.com to link the address to longitude and latitude coordinates, zip code and, most importantly, census tract and block group codes (see also Klick and Tabarrok, 2005).

Census block groups are the smallest geographical unit for which the bureau publishes sample data.[18] The census aims for a population of 1500 in each census block group. (In Washington, DC there are 433 census block groups with a mean population of 1321.) Using the census block group codes, one can merge geocoded crime data with data from the Census. Table 3.5, for example, shows two simple regressions of homicides and robberies on a handful of demographic variables by census block group.

The regressions should be considered exploratory: not only have we not tried to capture causality, but we have also not corrected for the spatial nature of the data, a subject we shall return to briefly further below. As explorations, the regressions indicate that homicides are more likely in census blocks with a large number of African Americans, less likely in blocks with Hispanics and less likely in high-income blocks. Of special interest is that homicides are much more likely in blocks with a large number of youths. Note in particular that the coefficient on youth is about ten times larger than the coefficient on African American.[19]

Robberies increase in blocks with more African Americans and also in blocks with more Hispanics, but although per capita income is of the same size as in the homicide regression it is no longer statistically significant. The coefficient on youths is also negative, of much smaller size and not statistically significant.

We can gain a deeper understanding of the data with the use of maps. The US Bureau of the Census (http://www.census.gov/) encodes geographic information in TIGER/Line→ files; 'TIGER' from the acronym **T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing. Conveniently, if the Census breaks data down by a type of region, it will typically also be possible to use TIGER/Line→ data to map the regions. TIGER/Line→ files, for example, encode geographic information such as boundaries for census block groups, census blocks, census tracts, counties, congressional districts, school districts and voting districts. TIGER/Line→ files can also include information on railroads, roads, landmarks and other geographic features.

Data from TIGER/Line→ files can be exported as 'shape files'. Shape files are a set of formats for geographic data developed by ESRI, a leading producer of software for spatial analysis. TIGER/Line→ shape files can be accessed from the US Bureau of the Census or more conveniently through ESRI, where Census 2000 TIGER/Line→ data can be downloaded by state[20] and according to a wide variety of layers such as census block groups, census blocks, census tracts, and so forth. We obtain boundary data on census block groups for Washington, DC in the form of shape files from ESRI.

Stata has no built-in functions for producing maps or using shape files. But two packages provide considerable functionality. SHP2DTA converts shape files to Stata dta files in a format that SPMAP can use to produce maps. Both packages can be easily installed in Stata using the ssc install command. The key to using these packages is to ensure that the location identifier in the map data is the same as the location identifier in the census data that you wish to map.[21]

Using these two packages, we produced Figure 3.9, a map of Washington, DC with homicides, robberies, and demographic data divided into eighths (15 percent quintiles) and shaded appropriately.

From the map we can see that homicides cluster in the south and east of Washington. Comparing with the demographic data, we can see that homicides occur in blocks with many African Americans, relatively few Hispanics, low per capita income and many youths. Robberies, in contrast, are located quite differently in the central areas of Washington. In particular, note the empty block in the southwest with zero African Americans, Hispanics, youths and no data on per capita income – this is the National Mall. Thus we can summarize the difference between homicides and robberies: homicides occur where the criminals are, robberies occur where the victims are.

Further inspection of the maps indicates the great separation in living space between African Americans in the southeast, Hispanics in the north central corridor and, by elimination, whites in the west. Rich and poor are equally divided, as indicated by the per capita income map, which of course also correlates with race. Finally, as noted above, homicides occur in blocks with many African Americans and many youths, but since these variables are highly correlated, more dispersion in the data is needed to convincingly disaggregate the influences of race, age and race combined with age.
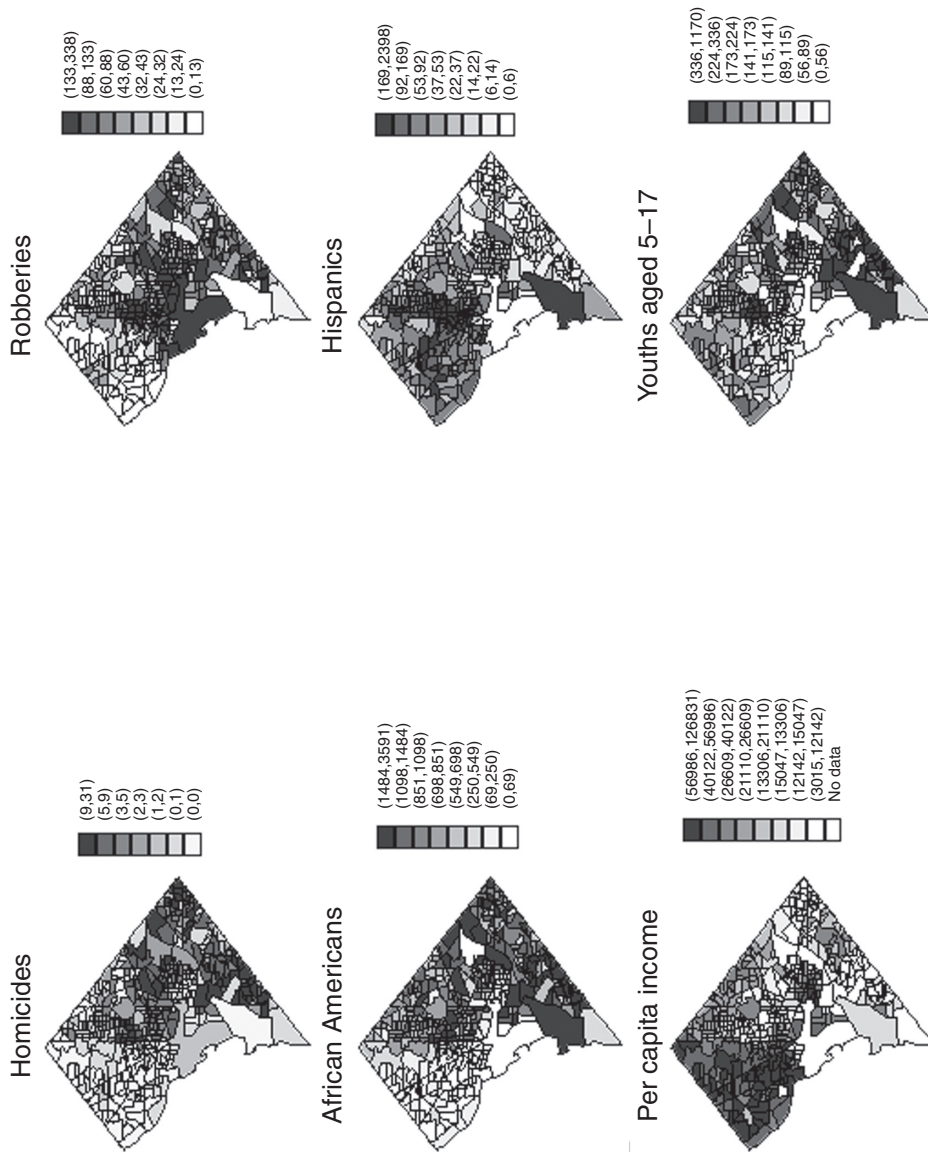
Robberies

(133,338)
(88,133)
(60,88)
(43,60)
(32,43)
(24,32)
(13,24)
(0,13)

Hispanics

(169,2398)
(92,169)
(53,92)
(37,53)
(22,37)
(14,22)
(6,14)
(0,6)

Youths aged 5–17

(336,1170)
(224,336)
(173,224)
(141,173)
(115,141)
(89,115)
(56,89)
(0,56)

Homicides

(9,31)
(5,9)
(3,5)
(2,3)
(1,2)
(0,1)
(0,0)

African Americans

(1484,3591)
(1098,1484)
(851,1098)
(698,851)
(549,698)
(250,549)
(69,250)
(0,69)

Per capita income

(56986,126831)
(40122,56986)
(26609,40122)
(21110,26609)
(13306,21110)
(15047,13306)
(12142,15047)
(3015,12142)
No data

Figure 3.9   Homicides, robberies and demographics in Washington, DC (January 1997–July 2003)

HOT SPOTS AND FUTURE DIRECTIONS

It has long been known that a handful of criminals account for a majority of crime. More recent research has shown that a handful of places account for a majority of crime. In Minneapolis, for example, 3 percent of street addresses and intersections generated 50 percent of all dispatched police calls (Sherman et al. 1989). By geocoding finer and finer data it is possible to identify factors that may account for such hot spots. For example, Sherman et al. (1989) find that five of the top ten hot spots in Minneapolis were located close to bars. In early work, Cohen (1980) found the obvious, that prostitution was prevalent in areas with large numbers of single males. But Cohen (1980) also found that street prostitution was more prevalent in areas where there were nearby unlit alleys, parks and parking lots.[22] The latter finding suggests how urban design, including such simple things as lighting, can also be considered a part of policing and crime control (Newman, 1973).

The analysis of spatial crime data has only scratched the surface. Cellular telephone records, directional GPS devices in automobiles, transponders used for road tolls – all capture geographic information that can and has been used in policing. In Massachusetts, some 700 people (as of 2008) are now required to wear electronic bracelets that track the wearers via satellite and alert the police if the wearer enters an 'exclusion zone' (Szep, 2008). Sex offenders, for example, are not permitted near schools, and restraining orders exclude abusers from being located near victims.

Proposals to track all offenders are now becoming common.[23] Before jumping too far on the GPS bandwagon, however, we need more research on the effectiveness of tracking. In an important paper, Agan (2011) used publicly available information on the home and work addresses of convicted sex offenders to see whether sexual assault locations correlated with the homes of convicted sex offenders. Interestingly, Agan found that, after controlling for other variables, the location of sexual assaults did not correlate well with the homes of convicted sex offenders – a useful caution before we jump too quickly towards 24-hour monitoring of all convicted criminals.

SPATIAL ECONOMETRICS

It is clear from the maps in Figure 3.9 that crime and its demographic correlates are not distributed randomly over space. Thus regressions like those in Table 3.5, which do not take into account spatial clustering or spatial autocorrelation, can be misleading. Spatial clustering is already familiar to most economists and corrections have been widely implemented since Moulton (1990). In a typical panel data situation, however, the geographic data become finer, and the clusters are no longer so obvious. There is no reason to believe, for example, that a census block group is the right cluster for city-wide crime data.

A spatial error model can take into account spatial clustering of errors due to omitted factors that span the data regions, in this case census block groups. A spatial lag model, analogous to temporal autocorrelation, recognizes that events in space may not be independent. An increase in crime in one region, for example, may predict an increase in crime in neighboring regimes through a variety of spillover processes. A gang of bank

robbers, for example, may create a crime wave in space, i.e. around certain locations, as well as through time. Gang wars ricochet across turf, i.e. space, generating correlation in homicide data; prostitution spills over from one block to another block and so forth.

In spatial econometrics, clusters and regions over which spatial autocorrelation may occur are defined by a weights matrix. The spatial lag model, for example, can be written (Anselin 1988):

$$Y = \rho Wy + X\beta + \varepsilon$$
$$\varepsilon \sim N(0, \sigma^2 I)$$

where W is a weights matrix with typical element w_{ij} indicating the 'distance' from region i to region j. Thus ρ can be thought of as the correlation between Y in one region and Y in that region's 'neighborhood'. Weights may be generated according to a wide variety of measures. Distance weights create for each geographic unit a distance measure to every other geographic unit, with units further away receiving a lower weight. Contiguity weights create for each geographic unit a binary variable equal to 1 if another geographic unit has a common border and zero otherwise. Higher-order contiguity weights can also be created that define two locations as 'connected' if they both border a third geographic unit. Weights may also be generated according to, say, an inverse square model, as with gravity models in trade analysis. Police are interested in finding 'hot pots' where crime is likely to be repeated. For this purpose, Bowers et al. (2004) define weights using a distance-time measure that weights geographically and temporally closer events as closer to one another. In short, weights are used to define 'neighborhoods' over which there are common influences. Neighborhood is a vague but necessary concept.

The spatial lag model cannot be estimated with OLS but can be estimated with maximum likelihood methods. The Stata package SG162 contains routines to create spatial weights, run spatial diagnostics and spatial regressions.[24] Using these routines, we re-ran the regressions in Table 3.5 but now using a spatial lag model that allows for autocorrelation across neighborhoods defined by a distance weight. The results are in Table 3.6.

For purposes of comparison, columns (1) and (3) contain the previous regressions results and (2) and (4) the results from the spatial lag model. The coefficients on the demographic variables, African Americans, Hispanics, and youths aged 5–17 are of similar size in any regression in which they were significant. The only notable change is that per capita income becomes less important and no longer statistically significant in the homicide regression, and it switches sign in the robbery regression (but is still not statistically significant). Notice that ρ – the measure of correlation between a region and its neighbor – is positive and statistically significant for both homicide and robberies and larger for robberies than for homicides.

What the spatial regression indicates is as follows. First, there is significant correlation between regions above and beyond that implied by similarities in demographic variables. Thus there is evidence for spillovers in crime processes. Second, we can see from the maps in Figure 3.9 that blocks with high per capita income tend to have low rates of crime and they tend to be far from blocks with large numbers of African Americans and youths. The ordinary regression, however, has no understanding of 'near' or 'far', and so ascribes much of the low crime to the high per capita income. The spatial lag

Table 3.6   Spatial lag models for homicides and robberies in Washington, DC

	(1) Homicides	(2) Homicides (spatial lag model)	(3) Robberies	(4) Robberies (spatial lag model)
Constant	1.962**	−0.295	33.854**	−12.790
	(0.523)	(0.619)	(10.217)	(10.835)
African	0.001*	0.001*	0.028**	0.023*
Americans	(0.001)	(0.001)	(0.011)	(0.010)
Hispanics	−0.001	0.0001	0.117**	0.087**
	(0.001)	(0.001)	(0.015)	(0.014)
Per capita	−0.048**	−0.018	−0.046	0.203
income (1000s)	(0.01)	(0.011)	(0.197)	(0.183)
Youths aged	0.013**	0.011**	−0.007	0.014
5–17	(0.002)	(0.002)	(0.037)	(0.034)
ρ		0.441***		0.650***
		(0.721)		(0.756)
Observations	429	429	429	429
R-squared	0.51	NA	0.21	NA

Notes:
Standard errors in parentheses.
 *   Significant at 5%.
**   Significant at 1%.
Data include all reports from January 1997 to July 2003.

model looks at the characteristics of each block as well as crime in neighboring blocks to estimate the influence of the independent variables. The fact that per capita income falls in importance once spatial lags are taken into account suggests that blocks with high per capita income that are near to blocks with large numbers of African Americans and youths have higher crime rates than would be expected based on per capita income alone. Thus the spatial regression recognizes that it is distance from high-crime regions, not high per capita income *per se*, that primarily reduces crime.[25] Although not statistically significant, the positive coefficient on robbery makes sense as it suggests that, if anything, criminals seeking financial gain prey on wealthier victims.

Spatial econometrics requires special techniques and is a rapidly growing field of research. We have only just touched on a few of the main issues. Interested readers may consult Ward and Gleditsch (2008) for an introductory treatment and Arbia and Baltagi (2008) for a variety of innovative applications.

CONCLUSION

Today we can measure vice and sin across time and space, we can map it correlate it and crunch it as never before. Moreover, the amount of crime data is growing in quantity and extent. Local crime data, including geographic information, for example, are exploding at a rapid pace. In fact, the amount of crime data can be overwhelming. We hope that

this chapter will help to guide researchers to the data that are most useful in answering their questions.

NOTES

1. We thank Adam Tabaka for research assistance. We thank Carlisle Moody, Justin McCrary and participants at the March 2009 DeVoe Moore Center Symposium on The Economics of Crime for comments.
2. Quoted in Benson and Hughes (1991, p. 112).
3. Even greater detail regarding circumstances surrounding homicides is available through the CDC's National Violent Death Reporting System (NVDRS). However, because the NVDRS began only in 2003, it is of less utility for researchers wishing to examine longer-term homicide patterns.
4. A density function is a 'sophisticated histogram.' The total area under the density function is equal to 1 and the area under the density function between any two points is the probability that a randomly drawn observation lies between those two points.
5. Aggravated assaults by definition do not include stolen property, which would convert an aggravated assault into a robbery. Arson reports are collected elsewhere.
6. For a list of exceptions to the rule see p. 19 of the FBI's *UCR Handbook* (2004), available online at http://www.fbi.gov/ucr/handbook/ucrhandbook04.pdf.
7. If there is not a reporting agency of similar population in the state, then a jurisdiction in the same region is used for estimation.
8. For more on imputation see Maltz and Targonski (2002) and Lynch and Jarvis (2008). For some other significant extrapolations, see Mosher et al. (2002, pp. 90–91).
9. It should also be noted that everything that has been said the offense data applies to an even greater extent to the arrest data (Maltz, 1999).
10. A list of the different NIBRS offense categories can be found at the FBI's NIBRS FAQ at http://www.fbi.gov/ucr/faqs.htm.
11. It should be noted that the victimization survey does not capture victimless crimes such as illicit drug use or prostitution; nor does it capture crime against business.
12. Most, but not all, files are available for use with Stata.
13. NCVS classifies the 40 largest metropolitan areas according to the total number of household interviews within each MSA.
14. Fowler (1993) offers a comprehensive guide to survey methods and problems.
15. In some instances those data extend beyond the three-year tracking period, although BJS makes no claims as to their completeness.
16. Data collection for a 2007 survey is complete but at the time of this writing these data have not yet been publically released.
17. Information about Stata at http://www.stata.com/. GeoDa is available from https://www.geoda.uiuc.edu/.
18. Census data are available from the US Bureau of the Census in many different forms. One of the most convenient is the online download tool, http://factfinder.census.gov/servlet/DownloadDatasetServlet?_lang=en.
19. In Washington, DC, youth and African American are highly correlated ($r = 0.84$), so we should not put too much weight on this finding although both coefficients are statistically significant.
20. http://www.esri.com/data/download/census2000_tigerline/index.html.
21. Useful information can be found here: http://www.stata.com/support/faqs/graphics/spmap.html.
22. Anselin et al. (2000) offer a useful survey of spatial crime analysis.
23. Potential victims are also subjecting themselves to tracking. Wealthy Mexicans are implanting GPS chips under their skin to help track them down in the event of a kidnapping (Daily_Mail_Reporter, 2008).
24. The Stata package is convenient but limited in important ways. In particular, a weights matrix for N regions must be N x N even if most of the elements are zero – this is memory intensive and Stata/IC is limited to 800 x 800 matrices. More extensive spatial procedures are available in R and Matlab. For the former see Bivand et al. (2008) and for the latter see LeSage and Pace (2009).
25. Note, however, that when a spatial lag is included, the interpretation of the other variables is different than with OLS. A change in variable X in district 1 has a direct influence on crime in district 1 and also influences crime in (potentially) every other district. In turn, crime in every other districts influences crime in district 1 through the spatial lag parameter. The coefficient in the regression measures the direct effect of a change in X and not the feedback effects.

REFERENCES

Agan, A. (2011), 'Sex offender registries: fear without function?', *Journal of Law and Economics*, forthcoming.

Alvazzi del Frate, A., O. Hatalak and U. Zvekić (eds) (2000), *Surveying Crime: A Global Perspective*, Rome: ISTAT.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Dordrecht, The Netherlands: Kluwer.

Anselin, L., J. Cohen, D. Cook, W. Gorr and G. Tita (2000), 'Spatial analyses of crime', *Criminal Justice*, **4**, 213–62.

Arbia, G. and B.H. Baltagi (2008), *Spatial Econometrics: Methods and Applications*, Heidelberg: Physica-Verlag.

Benson, D. and J. Hughes (1991), 'Method: evidence and inference – evidence and inference for ethnomethodology', in G. Button (ed.), *Ethnomethodology and the Human Sciences*, New York: Cambridge University Press: pp. 109–36.

Biderman, A.D. and D. Cantor (1984), 'A longitudinal analysis of bounding respondent conditioning and mobility as sources of panel bias in the National Crime Survey', in *Proceedings of the Survey Methods Research Section*, Alexandria, VA: American Statistical Association.

Biderman, A.D. and J.P. Lynch (1991), *Understanding Crime Incidence Statistics: Why the UCR Diverges from the NCS*, Research in criminology, New York: Springer-Verlag.

Bivand, R., E.J. Pebesma and V. Gómez-Rubio (2008), *Applied Spatial Data Analysis with R*, New York: Springer.

Blumstein, A. and J. Wallman (eds) (2000), *The Crime Drop in America*, Cambridge: Cambridge University Press.

Bowers, K., S.D. Johnson and K. Pease (2004), 'Prospective hot-spotting: the future of crime mapping?', *The British Journal of Criminology*, **44** (5), 641–58.

Cantor, D. and J.P. Lynch (2000), 'Self-report surveys as measures of crime and criminal victimization', in *Criminal Justice 2000*, vol. 4, Washington, DC: National Institute of Justice, pp. 85–138.

Catalano, S.M. (2007), 'Methodological change in the NCVS and the effect on convergence', in J. Lynch and L. Addington (eds), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, Cambridge Studies in Criminology, New York: Cambridge University Press, pp. 125–55.

Cohen, B. (1980), *Deviant Street Networks: Prostitution in New York City*, Lexington, MA: Lexington Books.

Daily_Mail_Reporter (2008), 'Terrified Mexicans splash out on chip implants so satellites can trace them if they're kidnapped', *The Daily Mail*, www.dailymail.co.uk/news/worldnews/article-1048129/Terrified-Mexicans-splash-chip-implants-satellites-trace-theyre-kidnapped.html.

Fowler, F.J. (1993), *Survey Research Methods*, Applied social research methods series (2nd edn), Newbury Park, CA: Sage Publications.

Freeman, R.B. (1996), 'Why do so many young American men commit crimes and what might we do about it?', *Journal of Economic Perspectives*, **10** (1), 25–42.

Fryer, R.G. Jr, P.S. Heaton, S.D. Levitt and K.M. Murphy (2005), 'Measuring the impact of crack cocaine', *SSRN eLibrary*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=720405, retrieved 13 May 2009.

Garicano, L. and P.S. Heaton (2006), 'Computing crime: information technology, police effectiveness and the organization of policing', *SSRN eLibrary*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=944874, retrieved 20 May 2009.

Grogger, J. (1998), 'Market wages and youth crime', *Journal of Labor Economics*, **16** (4), 756–91.

Groves, R.M. and D.L. Cork (2008), *Surveying Victims: Options for Conducting the National Crime Victimization Survey*, Washington, DC: National Academies Press.

Helland, E. and A. Tabarrok (2007), 'Does three strikes deter? A non-parametric investigation', *Journal of Human Resources*, **42** (2), 309–30.

Klick, J. and A. Tabarrok (2005), 'Using terror alert levels to estimate the effect of police on crime', *Journal of Law & Economics*, **48** (1), 267–79.

LeSage, J.P. and R.K. Pace (2009), *Introduction to Spatial Econometrics*, Boca Raton, FL: CRC Press.

Levitt, S.D. (2004), 'Understanding why crime fell in the 1990s: four factors that explain the decline and six that do not', *The Journal of Economic Perspectives*, **18** (1), 163–190. doi: 10.2307/3216880.

Lott, J.R. and J. Whitley (2003), 'Measurement error in county-level UCR data', *Journal of Quantitative Criminology*, **19** (2), 185–98. doi: 10.1023/A:1023054204615.

Lynch, P.L. (2006), 'Problems and promise of victimization surveys for cross-national research', *Crime and Justice*, **34**, 229–87.

Lynch, J.P. and J.P. Jarvis (2008), 'Missing data and imputation in the Uniform Crime Reports and the effects on national estimates', *Journal of Contemporary Criminal Justice*, **24** (1), 69–85. doi: 10.1177/1043986207313028.

Maltz, M.D. (1999), *Bridging Gaps in Police Crime Data: A Discussion Paper from the BJS Fellows Program*, Washington, DC: US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Maltz, M.D. and J. Targonski (2002), 'A note on the use of county-level UCR data', *Journal of Quantitative Criminology*, **18** (3), 297–318. doi: 10.1023/A:1016060020848.

Maltz, M.D. and J.Targonski (2003), 'Measurement and other errors in county-level UCR data: A Reply to Lott and Whitley', *Journal of Quantitative Criminology*, **19** (2), 199–206. doi: 10.1023/A:1023006321454.

Mayhew, H. (1861), *London Labour and the London Poor*, New York: A.M. Kelley.

Mayhew, P. and J. van Dijk (1997), *Criminal Victimisation in Eleven Industrialised Countries: Key Findings from the 1996 International Crime Victims Survey*, The Hague: Ministry of Justice, WODC.

McDowall, D. and C. Loftin (2007), 'What is convergence, and what do we know about it?', in J.P. Lynch and L.A. Addington (eds), *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, Cambridge Studies in Criminology, New York: Cambridge University Press, pp. 93–122.

Mosher, C.J., T.D. Miethe and D.M. Phillips (2002), *The Mismeasure of Crime*, Thousand Oaks, CA: Sage Publications.

Moulton, B.R. (1990), 'An illustration of a pitfall in estimating the effects of aggregate variables on micro units', *The Review of Economics and Statistics*, **72** (2), 334–8.

Newman, O. (1973), *Defensible Design*, New York: Macmillan.

O'Brien, R.M. (1985), *Crime and Victimization Data*, Law and criminal justice series, Beverly Hills, CA: Sage Publications.

Puzzenchera, C. and W. Kang (2008), 'Easy access to the FBI's Supplementary Homicide Reports: 1980–2005', online at http://ojjdp.nojrs.gov/ojstatbb/ezashr/.

Quetelet, A. (1831), *Research on the Propensity for Crime at Different Ages*, Cincinnati, OH: Anderson.

Sherman, L.W., P.R. Gartin and M.E. Buerger (1989), 'Hot spots of predatory crime: routine activities and the criminology of place', *Criminology*, **27** (1), 27–56.

Stamp, J. (1929), *Some Economic Factors in Modern Life*, London: P. S. King & Son Ltd. *Surveying Crime: A Global Perspective* (2000), Rome: ISTAT.

Szep, J. (2008), 'GPS grows as a crime-fighting tool in U.S', *Reuters*, http://features.us.reuters.com/techlife/news/19D02314-2148-11DD-86DE-177034A9.html.

Tabarrok, A. and E. Helland (2009), 'Measuring criminal spillovers: evidence from three strikes', *Review of Law & Economics*, **5** (1). doi: 10.2202/1555-5879.1326.

Turner, A.G. (1972), *San Jose Methods Test of Known Crime Victims*, Statistics technical report, Washington, DC: US Department of Justice.

US Department of Justice (2008), *Criminal Victimization in the United States, 2006 Statistical Tables*, Washington, DC: US Department of Justice.

Van Dijk, J. and P. Mayhew (1992), 'Criminal victimization in the industrialized world', The Hague: Directorate for Crime Prevention, Ministry of Justice, The Netherlands http://openlibrary.org/b/OL22241998M/Criminal-victimization-in-the-industrialized-world, retrieved 18 May 2009.

Van Kesteren, J.N., P. Mayhew and P. Nieuwbeerta (2000), 'Criminal victimisation in seventeen industrialised countries key findings from the 2000 International Crime Victims Survey', The Hague Ministry of Justice, WODC.

Ward, M.D. and K.S. Gleditsch (2008), *Spatial Regression Models*, Quantitative Applications in the Social Sciences, Los Angeles, CA: Sage Publications.

Witte, A.D. (1980), 'Estimating the economic model of crime with individual data', *The Quarterly Journal of Economics*, **94** (1), 57–84. doi: 10.2307/1884604.