

Measurement Error in Criminal Justice Data*

John Pepper
Department of Economics
University of Virginia
jvpepper@virginia.edu

Carol Petrie
Committee on Law and Justice
National Research Council
CPetrie@nas.edu

Sean Sullivan
Department of Economics
University of Virginia
sps2d@virginia.edu

June 28, 2009

Abstract

While accurate data are critical in understanding crime and assessing criminal justice policy, data on crime and illicit activities are invariably measured with error. In this chapter, we illustrate and evaluate several examples of measurement error in criminal justice data. Errors are evidently pervasive, systematic, frequently related to behaviors and policies of interest, and unlikely to conform to convenient textbook assumptions. Using both convolution and mixing models of the measurement error generating process, we demonstrate the effects of data error on identification and statistical inference. Even small amounts of data error can have considerable consequences. Throughout this chapter, we emphasize the value of auxiliary data and reasonable assumptions in achieving informative inferences, but caution against reliance on strong and untenable assumptions about the error generating process.

*We thank Stephen Bruestle, Alex Piquero, and David Weisburd for their helpful comments. Pepper's research was supported, in part, by the Bankard Fund for Political Economy.

1 Introduction

While accurate data are critical in understanding crime and assessing criminal justice policy, data on crime and illicit activities are invariably measured with error. Measurement errors occur to some degree in nearly all datasets, but are arguably more severe in surveys of illicit activities. Some individuals may be reluctant to admit that they engage in or have been victims of criminal behaviors, whereas others may brag about and exaggerate accounts of illicit activities. Administrative data on arrests or reported crimes are likewise susceptible to systematic recording errors and misreporting. The utility of data on illicit behavior is reduced when variables are measured with error.

In this chapter, we discuss the implications of measurement error for drawing inferences on crime and justice policy. We take as given that even the best-designed surveys of illicit activities are apt to suffer extensive and systematic data errors.¹ In light of this problem, we review the consequences of measurement error for identification and inference, and document key issues that should be addressed when using potentially mis-measured data.

We begin, in Section 2, with a brief review of several important measurement problems in data on crime and illicit behavior. We consider three related, but conceptually different, forms of data error: response error, proxy error, and imputation error. Response errors arise when variables of interest are observed, but possibly reported with error. This type of error is thought to be pervasive in self-report surveys on illicit activities, but can also be problematic in administrative data, such as the Uniform Crime Reports. Proxy errors arise when unobserved variables are replaced by related variables. An example is the use of the fraction of suicides committed with a firearm as a proxy measure for the rate of firearm ownership (Azrael et al., 2001). Finally, imputation errors arise when missing data are replaced by imputed values. A prominent example is the practice of imputing missing

¹An extensive literature attempts to document the validity and reliability of criminal justice data (Lynch and Addington, 2007; Mosher, Miethe and Phillips, 2002). In this chapter, we make no attempt to fully summarize this literature and offer no specific suggestions for modifying surveys to improve the quality of collected data.

observations in the Uniform Crime Reports.

In Sections 3 and 4, we formalize a number of statistical models to illustrate the impact of measurement error on inference. Repeatedly, we observe that data errors lead to fundamental identification problems, and therefore have important consequences for drawing informative inferences.

Section 3 concentrates on the case where mis-measured variables follow a convolution generating process, such that errors are approximated by additive unobserved random terms. Using a bivariate mean regression model, we first review the implications of the classical errors-in-variable model, where measurement errors are assumed to be mean zero and exogenous. In this classical model, the ordinary least squares estimator is downwardly inconsistent, and without additional data or assumptions, the mean regression can only be partially identified. Though a useful starting point, we argue that classical assumptions are frequently inappropriate for the study of crime and justice data, where errors are likely to be systematic and related to activities and policies of interest. When classical model assumptions are relaxed, the asymptotic properties of the ordinary least squares estimator cannot be easily characterized and, without additional data or assumptions, the mean regression is not identified.

Section 4 argues that for crime data, where outcomes are often discrete and only some observations are in error, a mixing model may be an appropriate framework for thinking about the effects of measurement error. Combined with fairly weak assumptions about the degree and nature of data errors, we show that the mixing model can be used to partially identify parameters of interest. Under weak assumptions, however, we observe that even small amounts of measurement error can lead to high degrees of ambiguity. Stronger assumptions on the unobserved error process lead to sharper inferences, but may not be credible in many cases.

Section 5 concludes with a discussion of the law of decreasing credibility (Manski, 2007). Agnostic models will often lead to indeterminate conclusions, while models imposing strong

and potentially inaccurate assumptions lead to less credible inferences. Throughout this chapter, we suggest situations in which apparent middle grounds may be available. In a variety of cases, one can draw informative inferences without imposing untenable assumptions.

2 Background and Evidence

The fundamental challenge of measuring crime is starkly illustrated by comparing different sources of crime data for the United States. The two most important sources are the Uniform Crime Reports (UCR) and the National Crime Victimization Survey (NCVS). For almost eight decades, the Federal Bureau of Investigation (FBI) has compiled the UCR by collecting information on arrests and crimes known to the police in local and state jurisdictions throughout the country. The NCVS, which began in 1973, is a general population survey conducted by the Bureau of Justice Statistics; it is designed to discover the extent, nature, and consequences of criminal victimization in the United States.

Table 1 displays a time-series of the rates (per 1,000) of rape, robbery, aggravated assault and property crime in the United States in 1990, 2000, and 2005, as reported in the official annual summaries of the UCR (U.S. Department of Justice, 2008) and NCVS (U.S. Department of Justice, 2006). Comparison of these two surveys reveals major differences in estimated crime rates and crime trends (see, for example, Blumstein and Rosenfeld, 2009; Lynch and Addington, 2007; McDowall and Loftin, 2007; NRC, 2003, 2008). Crime rates estimated from the NCVS are always substantially greater than those from the UCR. Although trends move in the same direction over this period, the estimated percentage drop in crime is notably more pronounced in the NCVS. For example, data from the UCR imply that annual rate of aggravated assaults fell from 4.2 in 1990 to 2.9 in 2005—a 31% drop—while data from the NCVS indicate that the annual rate fell from 9.8 in 1990 to 4.3 in 2005—a 56% drop.

[Table 1 about here.]

Such discrepancies are largely attributable to basic definitional and procedural differences between the two surveys (U.S. Department of Justice, 2004; Lynch and Addington, 2007). The two datasets measure different aspects of crime, with the UCR aiming to provide a measure of the number of crimes reported to law enforcement authorities, and the NCVS aiming to measure criminal victimization, including crimes not reported to authorities. These surveys also differ in the way they measure criminal behavior: the UCR is administrative data collected from individual criminal justice agencies (e.g. police departments), whereas the NCVS is a large-scale social survey that relies on self-reports of victimization.

Both methods of collecting data give rise to a number of response error concerns, including the potential for false reporting, non-standard definitions of events, and general difficulties associated with collecting information on sensitive topics and illegal behavior. In the NCVS, for example, self-reports by respondents who are concerned with the consequences of truthful admissions may yield a number of inaccurate reports. Likewise, police discretion in whether and how to record incidents may lead to substantial errors in the measurement of reported crimes in the UCR (Mosher, Miethe and Phillips, 2002: 84-86). In fact, Black (1970) observes that about one-quarter of reported felonies and just under one-half of reported misdemeanors are never formally recorded by the police.

In this section, we review several examples of measurement errors that are thought to confound inference on crime and criminal justice policy. Three related types of data errors are illustrated: response errors, proxy errors, and imputation errors.

2.1 Response Errors

Response errors arise when survey respondents misreport information. In the NCVS, for example, some respondents may misreport the incidence of crime; in the UCR, some police may fail to report or mis-classify reported crimes. Although there is much indirect evidence of response errors in survey data on illicit behavior, there is almost no direct evidence—especially for the basic crime data collected in the UCR and NCVS. An exception is a series

of validation studies that evaluate the incidence of illicit drug use by comparing self-reports to laboratory tests on hair, blood and urine. These validation studies suggest non-trivial and systematic self-reporting errors, but only apply to select populations (NRC, 2001).

2.1.1 Indirect Evidence on Response Errors

Indirect evidence on response error is often framed in terms of disparate findings between apparently similar surveys. A striking example is found in the literature examining the incidence of rape. A number of surveys reveal that between 20-25% of American women have been victims of completed or attempted rape at some point over their lifetime (Koss, 1993: table 1), yet NCVS data—which measure yearly, not lifetime, victimization—indicate that less than 0.1% of women experience a rape or attempted rape (Koss, 1996). Similarly divergent conclusions can be found in the most widely cited studies of the incidence of defensive gun use. Using data from the 1993 National Self-Defense Survey (NSDS), Kleck and Gertz (1995) estimate over 2 million defensive gun uses per year, yet the NCVS data from 1992 and 1994 reveals just over one-hundred thousand defensive gun uses per year (McDowall, et al., 1998).

These large differences have been attributed to the use of different survey questions, to sampling variability, and to response errors. As discussed in NRC (2005) and Tourangeau and McNeeley (2003), the surveys are structurally different, covering different populations, interviewing respondents by different methods, using different recall periods, and asking different questions. While the surveys attempt to study common topics, the particular measurements taken are different. Moreover, because rape and defensive gun use are sensitive, stigmatized and difficult to define, small differences in survey methods may lead to large differences in the quality of the self-reported data on these topics (NRC, 2003). Although response errors almost certainly affect these data, the extent of such errors is unknown.

2.1.2 Direct Evidence on Response Errors

In contrast to the previous examples, a number of validation studies provide direct evidence on the direction and magnitude of response errors in self-report data on illicit drug use. As summarized by NRC (2001) and Harrison and Hughes (1997), report-validation studies have been conducted on arrestees, addicts in treatment programs, employees, and persons in high-risk neighborhoods. Some of the most detailed and important validation studies were conducted with data from the Arrestee Drug Abuse Monitoring (ADAM)/Drug Use Forecasting (DUF) survey of arrestees, which elicits self-reports of drug use and also conducts urinalysis tests. Comparing self-reports of marijuana and cocaine use during the past three days to urinalysis tests for the same period, Harrison (1995) finds evidence of substantial and systematic response errors. In particular, she finds between 15 to 30 percent of respondents give inaccurate answers, with typically higher rates of false negatives than false positives. In the 1989 ADAM/DUF survey, for example, Harrison (1995) finds that 24.7 percent of respondents falsely deny using cocaine in the past three days, while 2.3 percent provide false positive reports. For marijuana use, 11.0% percent of responses are false negatives and 11.3% are false positives.

These validation studies provide some of the only direct evidence we have about the degree and nature of misreporting in surveys of illicit behaviors—we are not aware of similar direct evidence on response errors in other surveys of crime and victimization. Moreover, these studies provide only limited evidence about misreporting in national surveys on illicit drug use. The current validation studies examine particular subpopulations of individuals who have much higher rates of drug use than the general population. Response rates in the validation studies are often quite low, and respondents are usually not sampled randomly from a known population.

2.2 Proxy Variable Errors

In studies of illicit activities, it is often costly or impossible to directly measure variables of interest. In these cases, researchers may decide to use measurable variables as stand-ins or “proxies” for the unobserved variables. A proxy variable may be closely related to the unobserved variable of interest, but it is not a measurement of that variable. As a result, using a proxy in place of an unobserved variable can introduce substantial measurement error.

An important example is the use of UCR data on reported crime to act as a stand-in for the actual incidence of crime. A central but often overlooked problem is that many crimes are not reported to the police. The propensity of individuals to report crimes may be influenced by a variety of factors including the actual crime rate, the way victims are treated, community policing efforts, police manpower, and so forth (Rand and Rennison, 2002). It is therefore possible for crime reported to police, as measured by the UCR, to rise or fall independent of changes in actual crime. In this case, UCR data may result in misleading conclusions about crime trends and the impact of criminal justice policies on crime.

A number of other proxy variables are used in crime and justice research. For example, ecological studies evaluating the relationship between access to firearms and crime frequently rely on proxy variables for the rate of gun ownership (NRC, 2005). Proxies employed in the literature include the fraction of homicides committed with a firearm, the fraction of suicides committed with a firearm, and subscription rates to *Guns & Ammo* magazine (Azrael et al., 2001). Similarly, without access to information on the quantity of illicit drugs consumed, attempts to evaluate the demand for illicit drugs often rely on proxy variables. A common proxy for the actual quantity of drugs consumed is an indicator of whether an individual has used an illegal drug in a specified period of time. The accuracy of this proxy is unknown (NRC, 2001).

2.3 Imputation Errors

In nearly every survey, some members of the surveyed population choose not to respond to particular questions. This missing data problem is often addressed using imputation procedures that fill in missing data using values from complete records in the same dataset. Imputation errors arise when the values of imputed data differ from the true values of the underlying variable. Importantly, this measurement problem is conceptually different from the response and proxy variable problems. With imputation errors, the fraction of non-respondents is known, and the survey often identifies records with imputed values. This contrast with response and proxy variable problems, where the fraction of observations measured with error is generally unknown. Thus, in principle, we can learn much more about the implications of imputation error than response and proxy variable errors.

An important example of this type of error in criminal justice research involves the imputation of missing values in UCR data. These data are compiled through the voluntary reporting of local agencies. Some agencies neglect to provide information as called for in the reporting protocol, while others fail to report altogether (Maltz, 1999; Maltz and Targonski, 2002). In 2003, for example, over one-third of agencies filed incomplete monthly reports (Lynch and Jarvis, 2008). A higher frequency of non-responding agencies come from small jurisdictions, so that the agencies with missing data serve just over twelve-percent of the population.

The FBI uses a simple procedure to impute the values of crime data for non-responding agencies. Missing values for agencies reporting three or more months of crime data are replaced by the agency-average crime rate over observed months. For agencies reporting two or fewer months, crime rates from agencies of similar size and location are used (for details, see Maltz, 1999).

Surprisingly little research has been aimed at examining the inferential implications of imputed response problems in the UCR. Maltz and Targonski (2002) argue that imputation methods are likely to bias estimates of crime rates and program evaluation at the county

level, but may be benign for analyses at higher geographic aggregations. Lynch and Jarvis (2008), however, find that the imputations can have a large impact on estimates of national trends in crime: from 1992 to 1993, UCR data without imputed values indicate a 4.7% drop in the volume of offenses, whereas this drop is only 1.9% when imputed values are included in the data.

3 The Convolution Model

Because data used to monitor crime and evaluate crime policy are invariably measured with error, it is important to understand how such errors may impact inferences, and whether anything can be done to credibly mitigate negative effects. Exactly how data errors affect statistical inference depends heavily on the specifics of the problem: critical details include the type of model being estimated, the way measurement errors enter the model, and the joint distribution of the observed and unobserved random variables.

In this section, we use a convolution model, where errors are approximated by unobserved additive random terms, to focus on the problem of drawing inferences on a bivariate mean regression model. This model of data errors can be used to formalize the effects of response errors, and with minor modifications can also be generalized to accommodate proxy variable errors. Imputation errors, where only a known fraction of observations may be in error, are probably better addressed by the mixing model considered in Section 4.

Consider the bivariate mean regression model

$$y^* = \alpha + x^* \beta + \varepsilon \tag{1}$$

where y^* and x^* are scalars and ε is an unobserved random variable distributed according to some probability distribution F_ε , mean independent of x^* : formally, we assume

$$[\text{A1}] \ E[\varepsilon|x^*] = 0$$

Assumption A1 is the standard mean independence requirement for linear regression models. Given a random sample on (y^*, x^*) , one can consistently estimate (α, β) using the ordinary least-squares estimator.

In the presence of data errors, however, x^* and y^* may not be revealed by the sampling process. Rather, the convolution error model assumes observable data are imperfect reflections of the true variables of interest: $x = x^* + \mu$ and $y = y^* + \nu$, with unobserved measurement errors, μ and ν , randomly distributed according to probability distributions F_μ and F_ν respectively. The sampling process does not reveal the joint distribution of the variables of interest, (x^*, y^*) , but it does reveal that of (x, y) . Specifically, we assume an observable random sample of size N : $\{(x_i, y_i)\}_{i=1}^N$.²

What can this sampling process reveal about the parameters of interest, namely (α, β) ? Under various assumptions on the characteristics of the error distributions, F_μ and F_ν , we explore the effects of measurement error on the probability limit of the least squares slope estimator, and suggests potential solutions to the inferential problems caused by data errors. We first review the classical errors-in-variables model, and then motivate and consider several non-classical variations on this model. Finally, we discuss complications introduced by measurement error in regressions on differenced panel data. These examples are by no means exhaustive, but help give a taste for important considerations and implications when using error-ridden measurements in statistical analysis.³ In particular, we observe the following:

- (i) the “classical” generating processes, where errors are assumed to be mean zero and exogenous, is unlikely to apply in many of the important data error problems in criminal justice research;

²To accommodate proxy errors, this model has often been generalized by including a factor of proportionality linking the observed and true variables. For example, $y = \delta y^* + \nu$, where δ is an unknown parameter. For brevity, we focus on the pure measurement model without an unknown factor of proportionality. Including a scaling factor of unknown magnitude or sign induces obvious complications beyond those discussed here. For additional details, see Wooldridge (2002: 63-67) and Bound et al. (2001: 3715-3716).

³The interested reader should consult more detailed presentations in Wooldridge (2002), Wansbeek and Meijer (2000) and Bound, Brown, and Mathiowetz (2001).

- (ii) outcome-variable errors, ν , impact inferences on the mean regression in Equation (1) and, in particular, can bias inferences on crime levels and trends;
- (iii) regressor errors, μ , do not always result in attenuation bias; and
- (iv) in panel data models, the impact of measurement error can be exaggerated by differencing or the inclusion of fixed effect terms.

In the following, let $\sigma_{x^*}^2$ denote the population variance of x^* , with similar notation for other variables, and let $\sigma_{x^*,\mu}$ and $\rho_{x^*,\mu}$ denote the population covariance and correlation of x^* and μ with similar notation for other pairwise combinations of variables. In regression models, we adopt the terminology that y^* is the “outcome variable” while x^* is the “regressor.” Finally, let $\hat{\beta}_{y,x}$ be the ordinary least squares (OLS) slope estimator from a sample regression of y on x .

3.1 Classical Assumptions

The classical measurement error model supposes that the additive error terms μ and ν are mean zero, uncorrelated with the true values of all variables in the model, and uncorrelated with each other. In the present model, classical measurement error assumptions may be stated as follows:

$$[\text{A2}] \quad E[\mu] = E[\nu] = 0$$

$$[\text{A3}] \quad \sigma_{x^*,\nu} = \sigma_{\mu,\nu} = \sigma_{\varepsilon,\nu} = 0$$

$$[\text{A4}] \quad \sigma_{x^*,\mu} = 0$$

$$[\text{A5}] \quad \sigma_{\varepsilon,\mu} = 0.$$

Assumption A2 implies that the measurement errors, μ and ν , are mean zero, while the remaining assumptions restrict errors to be uncorrelated with each other, with the outcome variable, and with the regressor. In particular, assumption A3 implies that the error in

the outcome variable, ν , is uncorrelated with the other measurement error, μ , with the regressor x^* , and with the outcome variable y^* . Assumptions A4 and A5 restrict the error in the regressor, μ , to be uncorrelated with the true value of the regressor, x^* , and with the outcome, y^* .

Under A1-A5, it is well known that the probability limit of the OLS slope parameter is proportional to the true value β in the following way (Wooldridge, 2002):

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{y,x} = \beta \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_{\mu}^2}. \quad (2)$$

Two important conclusions may be drawn from Equation (2). First, measurement error in the outcome variable does not affect the consistency of $\hat{\beta}_{y,x}$: the slope coefficient from a sample regression of y on x^* is a consistent estimator of β .⁴ Second, classic measurement error in the regressor causes the sample regression slope parameter to be an inconsistent estimator of β such that asymptotically $\hat{\beta}_{y,x}$ has the same sign as β but is closer to zero. This effect is generally termed “attenuation bias.” The presence of measurement error in the regressor dilutes the apparent strength of the relationship between x^* and y^* , causing the estimated slope parameter to understate the magnitude of the true effect. While we focus on the OLS estimator of the slope parameter, the estimator of the constant term, α , is also inconsistent when the regressor is measured with error.⁵

With access to auxiliary data or model structure, the parameters (α, β) may be point identified.⁶ For example, one common approach is to exploit an instrumental variable, z , that is known to be independent of all of the unobserved error terms, (μ, ν, ε) , but is

⁴When the available measurement of y^* is a proxy variable of the form $y = \delta y^* + \nu$, the probability limit of $\hat{\beta}_{y,x^*}$ is $\delta\beta$. If δ is known in sign but not magnitude, then the sign but not scale of β is identified.

⁵Although a full treatment of the effects of measurement error in multivariate regression is beyond the scope of this chapter, several general results are worth mentioning. First, measurement error in any one regressors will usually affect the consistency of all other parameter estimators. Second, when only a single regressor is measured with classical error, the OLS estimator of the coefficient associated with the error-ridden variable suffers attenuation bias in the standard sense (see, for example, Wooldridge, 2002: 75). In general, all other OLS parameters are also inconsistent, and the direction of inconsistency can be asymptotically signed by the available data. Finally, with measurement error in multiple regressors, classical assumptions imply the probability limit of the OLS parameter vector is usually attenuated in an average sense, but there are important exceptions (Wansbeek and Meijer, 2000: 17-20).

⁶A number of possible strategies are available, and the interested reader should consult the discussions in Wansbeek and Meijer (2000) and Bound et al. (2001).

also correlated with the regressor, x^* . In particular, assume that

$$[A6] \sigma_{z,x^*} \neq 0$$

$$[A7] \sigma_{z,\mu} = \sigma_{z,\nu} = 0$$

$$[A8] \sigma_{z,\varepsilon} = 0$$

In this case, the instrumental variable (IV) estimator is consistent for β :

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{y,x(z)}^{IV} = \beta. \quad (3)$$

It is often observed that alternative measurements of x^* may serve as instrumental variables satisfying these conditions. For example, suppose a researcher has access to an alternative measurement $x' = x^* + \eta$ with η randomly distributed according to some probability distribution F_η . If error in the alternative measurement, η , is classical and uncorrelated with μ , then $z = x'$ satisfies A6-A8 and a consistent point estimator of β is available even in the presence of classical measurement error.⁷

In the absence of auxiliary data or structure, point identification of β is impossible. Under classical assumptions, however, the sampling process places bounds on the true value of β (Frisch, 1934). Part of the work is already done: with classical measurement error in the outcome variable, equation (2) shows that probability limit of $\hat{\beta}_{y,x}$ is closer to zero than is the true value of β , so that $|\hat{\beta}_{y,x}|$ is an estimable lower bound on the magnitude of β .

Now consider a new estimator of β constructed from the reverse regression of x on y . Specifically, define the new estimator of β as the inverse of the ordinary least squares estimator of the slope from a regression of x on y : $\hat{\beta}_{x,y}^{-1}$. Under A1-A5, the probability limit

⁷Of the required conditions for using a second measurement as an instrumental variable, the assumption that the two errors are uncorrelated, $\sigma_{\eta,\mu} = 0$, may be the most difficult to satisfy in practice. Even if both errors are classical in other regards, errors in different measurements of the same variable may be expected to correlate so that $\sigma_{\eta,\mu} > 0$. When covariance between μ and η is non-zero, the IV slope estimator is no longer a consistent point estimator of β , though it may still provide an informative bound on β under certain circumstances (see, for example, Bound et al., 2001: 3730; Black et al., 2000).

of this new estimator is

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{x,y}^{-1} = \beta + \frac{\sigma_\varepsilon^2 + \sigma_\nu^2}{\beta \sigma_{x^*}^2} \quad (4)$$

Like the usual slope estimator from a regression of y on x , the probability limit of $\hat{\beta}_{x,y}^{-1}$ has the same sign as β . Unlike the usual slope estimator, however, the probability limit of $\hat{\beta}_{y,x}$ is farther from zero than is the true value of β .

This means that in the presence of classical measurement error, data on x and y alone can be used to provide informative bounds on the set of possible values of β . When $\beta > 0$, $\hat{\beta}_{x,y}^{-1}$ and $\hat{\beta}_{y,x}$ are asymptotic upper and lower bounds on the true value of the slope parameter:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{y,x} < \beta < \text{plim}_{N \rightarrow \infty} \hat{\beta}_{x,y}^{-1} \quad (5)$$

and inequalities reverse when $\beta < 0$.⁸

3.2 Problems with Classical Assumptions

In any analysis with imperfectly measured data, the researcher should carefully consider whether classical assumptions are appropriate. Trivial examples which gives rise to classical conditions are the cases where errors are generated by random clerical mistakes in data entry or by sampling variation when x^* and y^* represent population averages (Bound et al. 2001). In general, however, classical measurement error assumptions are inappropriate. Failure of any one of the assumptions has implications for drawing inferences on the regression parameters (α, β) .

To evaluate these problems, we present two illustrations where non-classical measurement errors are likely to confound inference: these are the common monitoring problem of inferring levels and trends in crime rates, and the problem of inferring how expected crime rates vary with illicit drug use. In each situation, we argue that various classical assumptions are unlikely to hold. We then derive formal results to illustrate the implications of alternative assumptions on the measurement error process.

⁸Klepper and Leamer (1984) suggest a similar strategy for the case where multiple regressors are measured with error. The general approach is described by Bound et al. (2001: 3722-3723).

3.2.1 Illustration 1: Crime Levels and Trends

Perhaps the most central function of the data on crime is to monitor levels and trends in crime rates. Level estimates from the UCR and NCVS, however, are thought to be downwardly biased. This violates assumption A2, because measurement errors are not mean-zero.

As estimated crime rates are perceived to be systematically biased, trend estimates are often argued to be more reliable than levels. An example is the study of differences in crime statistics between the UCR and NCVS; these two crime series differ systematically in levels, with the NCVS always estimating a higher rate of crime than the UCR (see Table 1). Despite obvious level differences, some researchers suggest that the two series are comparable in terms of long-term time trends (U.S. Department of Justice, 2004). On this topic, McDowall and Loftin (2007: 96) argue that studies attempting to reconcile differences between the UCR and NCVS crime rates might want to focus on differences in trends as a less-demanding standard of comparability than differences in levels.

Similar sentiments have been expressed in the context of survey data on illicit drug use. To the extent that survey respondents may be reluctant to admit engaging in illegal and socially unacceptable behavior, it seems likely that drug use data may suffer from a systematic, negative bias. Based on the premise that measurement errors are constant over time, Johnston et al. (1998: 47-48) argue that measurements of drug use trends should be robust to the presence of measurement error. Anglin et al. (1993: 350) take a similar stance, claiming “[I]t is easier to generate trend information... than to determine the absolute level.”

Under what conditions might one consistently estimate trends, even if the level estimates are systematically biased? Let x^* be an indicator function measuring two distinct time periods, say 2009 and 2010, and suppose the noisy reflection of the crime rate satisfies $y = y^* + \nu$. Assume that the measurement error, ν , is uncorrelated with the time period so that $\sigma_{x^*,\nu} = 0$, but allow for the possibility of non-zero expected errors (e.g., $E[\nu] < 0$). That is, maintain assumption A3 but relax A2 to allow for non-zero mean errors. In this case, with only the outcome variable measured with error, the OLS estimator of the slope

parameter from a regression of y on x^* is a consistent estimator of β (see Equation 2) but the estimator of α is generally inconsistent:

$$\text{plim}_{N \rightarrow \infty} \hat{\alpha}_{y, x^*} = \alpha + E[\nu] \quad (6)$$

Thus, even when the available measurement of y^* is systematically biased, we can consistently estimate the trend, β , but not the level of $E[y^*|x^*]$.

The result that trends are robust to biased measurements of the outcome variable is critically dependent on the assumption that measurement error is uncorrelated with the observed regressor: assumption A3. The reason results depend so heavily on this condition is easily seen in terms of general conditional expectations of y at arbitrary values $x^* = x_a^*$ and $x^* = x_b^*$. For any given value of x^* —say, $x^* = x_a^*$ —the expected value of y is

$$E[y|x^* = x_a^*] = E[y^*|x^* = x_a^*] + E[\nu|x^* = x_a^*] \quad (7)$$

and the difference in conditional expectations is

$$\begin{aligned} E[y|x^* = x_a^*] - E[y|x^* = x_b^*] &= (E[y^*|x^* = x_a^*] - E[y^*|x^* = x_b^*]) \\ &\quad + (E[\nu|x^* = x_a^*] - E[\nu|x^* = x_b^*]) \end{aligned} \quad (8)$$

Thus, the observed level is equal to the true level only when $E[\nu|x^*] = 0$, which is violated in this example. The observed difference, on the other hand, is equal to the true difference under the weaker condition that $E[\nu|x^*] = E[\nu]$ for all x^* . Intuitively, if the expected measurement error is mean independent of the conditioning variable, then the bias terms cancel out so that changes in conditional expectations of y are the same as changes in the conditional expectation of y^* .

As previously asserted, the critical assumption is that measurement errors in y are unrelated to the conditioning variable x^* . This assumption is particularly questionable when x^* represents time. Consider measurement errors in self reports of illicit drug use: changes over time in the social and legal stigma of drug use seem likely to correlate with response errors (see Pepper, 2001). Similarly for crime rate data in the UCR, the frequency and quality of

reporting seems likely to vary over time in response to changes in laws, social attention to crime, and the crime rate itself (Mosher, Miethe and Phillips, 2002; Rand and Rennison, 2002).

Non-classical measurement errors in the outcome variable are also likely to impact numerous program evaluation studies of criminal justice policy. Consider, for example, the problem of assessing the impact of the police-force size or policing practices on crime, or the impact of restrictive firearms policies on crime. In these cases and many others, errors in measuring crime are likely to be associated with effects of the policies of interest.

When measurement errors in crime data are associated with the conditioning variable, analysis in trends or marginal changes is not clearly preferable to that in levels. Comparing bias terms in Equations (7) and (8), the absolute magnitude of $E[\nu|x^*]$ may be greater than that of $E[\nu|x_a^*] - E[\nu|x_b^*]$, but relative to $E[y^*|x^*]$ and $E[y^*|x_a^*] - E[y^*|x_b^*]$, the impact of the bias term in changes may well exceed that in levels. Certainly, the claim that crime data are biased in levels but not trends cannot generally be supported.

3.2.2 Illustration 2: Drugs and Crime

In many cases, concerns over errors in the regressor may also play an important role in inference. As an example, suppose a researcher sought to measure the effect of drug use on the propensity to engage in criminal behavior (Bennett et al., 2008; Chaiken and Chaiken, 1990). Assume the bivariate mean regression model in Equation (1), where x and y are indicator functions for self reports of drug use and criminal activity, respectively. To properly interpret the slope parameter from a sample regression of y on x , a detailed understanding of the sources of measurement error is required. Far from random clerical errors or sampling variability, measurement errors in this model potentially violate classical assumptions A3-A5.

Take assumption A3, which requires measurement error in the outcome variable to have zero covariance with both the true value of the regressor and the measurement error in this variable. If individuals who falsely deny engaging in one of the activities are likely to falsely deny engaging in both (with similar logic for false positives), then measurement errors in self

reports of x^* and y^* will positively correlate in the population: $\sigma_{\nu,\mu} > 0$. Likewise, errors in reporting criminal activity, ν , are arguably related to whether the respondent actually used illicit drugs, x^* .

Because x^* is a binary variable in this model, assumption A4 is violated by definition. To see why, note that when $x^* = 0$ any error must be a false positive, and when $x^* = 1$ any error must be a false negative. As such, errors in the regressor exhibit negative correlation with true values of the variable, violating assumption A4.⁹ Of course assumption A4 can also be violated in cases where x^* is not binary. Suppose binary reports were replaced by state-level aggregates of self-reported drug use and criminal behavior. If the negative stigma of drug use decreases as drug use becomes more mainstream, then errors in the state-level statistics should again negatively correlate with actual aggregate drug use: $\sigma_{x^*,\mu} < 0$.

Finally, assumption A5 fails when measurement error in the regressor is correlated with unobserved random variable, ϵ , in the population regression. An active police presence in the community may tend to reduce the likelihood that an individual would actually commit a crime and at the same time make individuals less likely to truthfully admit illicit drug use, leading to negative covariance between measurement error in the regressor, μ , and the regression error term, ϵ . A similar story can be told at the state level. Violence and other criminal behavior in areas with higher-than-average crime rates may desensitize these populations to the negative stigma associated with admitting drug use, again leading to negative covariance between drug use statistics and the error term in the population model: $\sigma_{\mu,\epsilon} < 0$.

3.3 Non-Classical Assumptions

As the previous illustrations demonstrate, classical assumptions may be invalid in many important applications. Although the effects of measurement error on the consistency of the OLS slope estimator were simple to characterize under classical errors-in-variables assump-

⁹Similar logic suggests violation for any discrete variable, and any continuous but bounded variable.

tions (A2-A5), relaxing these assumptions leads to a less parsimonious probability limit:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{y,x} = \frac{\sigma_{x^*,\nu} + \sigma_{\mu,\nu}}{\sigma_{x^*}^2 + \sigma_{\mu}^2 + 2\sigma_{x^*,\mu}} + \beta \frac{\sigma_{x^*}^2 + \sigma_{x^*,\mu}}{\sigma_{x^*}^2 + \sigma_{\mu}^2 + 2\sigma_{x^*,\mu}} + \frac{\sigma_{\mu,\varepsilon}}{\sigma_{x^*}^2 + \sigma_{\mu}^2 + 2\sigma_{x^*,\mu}} \quad (9)$$

Each term on the right-hand side of (9) corresponds to the failure of one of the classical assumptions: A3-A5.¹⁰ The first term is non-zero when errors in the outcome are related to the true value of the regressor or its measurement error, so that assumption A3 is not satisfied. The second term differs from the classical result when errors in the regressor are related to the true value of the regressor, $\sigma_{x^*,\mu} \neq 0$, so that assumption A4 is not satisfied. The third term is non-zero when the errors in the regressor are related to the regression residual, so that assumption A5 does not hold.

In this more general setting, the two central lessons of the classical errors-in-variables model no longer apply: data errors in the outcome variable have consequences for inference, and measurement errors in the regressor need not bias the OLS estimator towards zero. In particular, because the first term in Equation (9) is unrestricted in both size and sign, the failure of A3 alone is a sufficient condition for $\hat{\beta}_{y,x}$ to be potentially inconsistent for both the sign and magnitude of the true slope parameter. Likewise, the failure of assumption A5—where the error in the regressor, μ , is related to the regression error, ε —leads to an additive term which is unrestricted in size and sign. Failure of A5 is also sufficient to cause general inconsistency of the OLS estimator of β .

Even if A3 and A5 hold, so that the only deviation from classical assumptions is non-zero covariance between the true value of the regressor and its measurement error, attenuation bias is not guaranteed. If $\sigma_{x^*,\mu} > 0$, then $\hat{\beta}_{y^*,x}$ is consistent for the sign of β and indeed suffers from attenuation bias. However, when $\sigma_{x^*,\mu} < 0$, as might be the case in a study of the impact of illicit drug use on crime, non-classical error in the regressor may lead to arbitrary inconsistencies. Depending on the relative variance and correlation of x^* and μ , the probability limit of the OLS slope estimator may have the incorrect sign, or may artificially amplify (rather than attenuate) the strength of the relationship between the outcome variable

¹⁰Failure of assumption A2 affects inference regarding α , but not β (see, for example, Illustration 1).

and regressor.

As should be clear at this point, non-classical measurement error is a more insidious concern than classical error. Without auxiliary data or structure, the sampling process places no restrictions on the true value of β , and the prospects for identification and estimation of β are grim.

Under certain restrictions, however, it may still be possible to construct informative bounds on β . For example, consider the situation where the only deviation from classical assumptions is a negative covariance between the regressor and its measurement error; $\sigma_{x^*,\mu} < 0$. Thus, assumptions A2, A3 and A5 are assumed to hold, but assumption A4 does not. Assume for simplicity that $\beta > 0$. As long as x^* and μ are not too highly correlated, the OLS estimator still acts as an asymptotic lower bound on β , and under fairly weak conditions the IV estimator using an alternative measurement of x^* is an asymptotic upper bound on the true slope parameter (Black et al., 2000).¹¹ Thus, the value of β can be bounded in a manner similar to the case for the classical errors-in-variables model:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{y,x} < \beta < \text{plim}_{N \rightarrow \infty} \hat{\beta}_{y,x(x')}^{IV} \quad (10)$$

and inequalities reverse when $\beta < 0$.

3.4 Panel Data Models

Up to this point, we have maintained assumption A1 so that OLS estimators of (α, β) would be consistent if x^* and y^* were observable in the data. In many cases, however, this assumption is likely to be violated. A common example is the situation where multiple observations are collected on each sampling unit: for example, let i index a county and let t index the year of the observation. Let the unobserved regression error contain unit-specific effects, α_i , so that $\varepsilon_{i,t} = \alpha_i + \omega_{i,t}$ where α_i is potentially correlated with $x_{i,t}^*$ and $\omega_{i,t}$ is an unobserved random variable distributed according to some probability distribution F_ω , mean

¹¹Bollinger (1996) and Frazis and Lowenstein (2003) derive bounds when a binary regressor is measured with error.

independent of x^* . To simplify discussion, suppose $y_{i,t}^*$ is measured accurately, $x_{i,t} = x_{i,t}^* + \mu_{i,t}$ is an imperfect measurement of $x_{i,t}^*$, classical measurement error assumptions hold, and all variables have stationary variance and covariance across counties and time.¹²

We assume access to a panel-data sample of two years worth of data: $\{(x_{i,t}, y_{i,t})\}_{t=1}^2\}_{i=1}^N$. Denote by $\hat{\beta}_{y,x}$ the OLS slope estimator from regression of $y_{i,t}$ on $x_{i,t}$ which does not account for variation in α across counties. As should be expected, $\hat{\beta}_{y,x}$ is an inconsistent estimator of β :

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{y,x} = \beta \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_{\mu}^2} + \frac{\sigma_{x^*,\alpha} + \sigma_{\mu,\alpha}}{\sigma_{x^*}^2 + \sigma_{\mu}^2} \quad (11)$$

The first term in Equation (11) is just the classical attenuation bias observed in the linear regression models without unobserved effects. The second term comes from failure to account for unobserved county-specific effects: these terms end up in a composite regression-error term which correlates with x_i^* , effectively violating the assumption of conditional mean-zero regression errors (assumption A1) and resulting in an additional source of inconsistency.

To avoid problems caused by ignoring time-constant unobserved effects, the researcher may exploit the panel structure of the collected data. For example, a first difference (FD) estimator eliminates any time-constant terms while leaving β estimable:

$$\Delta y_i^* = \Delta x_i^* \beta + \Delta \varepsilon_i \quad (12)$$

where $\Delta y_i^* = y_{i,2}^* - y_{i,1}^*$ with similar notation for other variables, and where $\Delta \alpha_i = 0$ by definition. With an accurate measurement of x_i^* available to the researcher, OLS performed on Δy_i^* and Δx_i^* would provide a consistent estimator of β .

Since x_i^* is not observed, let $\hat{\beta}_{\Delta y^*, \Delta x}$ denote the slope parameter from a sample regression of Δy_i^* on Δx_i . Under classical assumptions, $\hat{\beta}_{\Delta y^*, \Delta x}$ suffers from standard attenuation bias:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{\Delta y^*, \Delta x} = \beta \frac{\sigma_{\Delta x^*}^2}{\sigma_{\Delta x^*}^2 + \sigma_{\Delta \mu}^2} \quad (13)$$

At first glance, Equation (13) would seem an improvement upon Equation (11), with consistency of $\hat{\beta}_{\Delta y^*, \Delta x}$ only limited by attenuation bias due to the presence of measurement error

¹²With panel data, assumptions A1-A5 must account for correlations in both the cross-section and the time series dimension. For detailed examples, see Wooldridge (2002) and Griliches and Hausman (1985).

in x_i . While the FD estimator does eliminate inconsistency due to the presence of time-constant unobserved variables, it may also exacerbate attenuation bias due to measurement error. More extreme attenuation bias results if the true regressor exhibits relatively strong serial correlation while measurement error does not.¹³

Switching from levels to differences is not uniformly preferable when the regressor is measured with error. On one hand, working with differences allows the researcher to eliminate inconsistency due to the presence of time-constant unobserved effects. On the other hand, differencing and related approaches may tend to increase the magnitude of measurement error bias when true values of the regressor exhibit strong serial correlation and measurement errors do not. For a detailed discussion of related panel data models and solutions when more than two periods of data are available, see Griliches and Hausman (1985).

4 The Mixture Model

The previous section presented textbook results for several chronic errors-in-variables models, where the observable variable, y , is the noisy reflection of the true variable of interest, y^* , such that $y = y^* + \nu$. In many settings, this model of chronic errors may be inappropriate. When considering imputation errors in the UCR, for example, we know that some observations are imputed while others are not. Likewise, when focusing on invalid response problems regarding victimization in the NCVS or illicit drug use in the National Survey of Drug Use and Health (NSDUH) it seems likely that some respondents report accurately while others may not.

In light of these concerns, a growing body of literature conceptualizes the data error problem using a mixture model in which the observed outcome distribution is a mixture of the unobserved distribution of interest, and another unobserved distribution. See, for example, Horowitz and Manski (1995), Lambert and Tierney (1997), Dominitz and Sherman

¹³Note that the variance of Δx_i^* is smaller when x_i^* has positive autocorrelation: $\sigma_{\Delta x^*}^2 = 2\sigma_{x^*}^2(1 - \rho_{x_2^*, x_1^*})$. To see why the relative strength of serial correlation is a concern, suppose that $\rho_{x_2^*, x_1^*} > 1/2$, while random measurement errors exhibit no autocorrelation. This implies $\sigma_{\Delta x^*}^2 < \sigma_{x^*}^2$ while $\sigma_{\Delta \mu}^2 = 2\sigma_{\mu}^2$, so attenuation bias will be greater after first differencing the data.

(2004), Mullin (2005) and Kreider and Pepper (2007, 2008, forthcoming). In this setting, the observed random variable, y , is viewed as a contaminated version of the variable of interest, y^* . In particular, the observed variable is generated by the mixture:

$$y = y^* z + \tilde{y}^*(1 - z) \tag{14}$$

where z indicates whether the observed outcome, y , comes from the distribution F_{y^*} or some alternative distribution, $F_{\tilde{y}^*}$.

In this environment, the “contaminated sampling” model pertains to the case in which the mixing process, z , is known to be statistically independent of sample realizations from the distribution of interest, y^* . The more general “corrupted sampling” model pertains to the case where nothing is known about the pattern of data errors.

Using nonparametric methods, Horowitz and Manski (1995) derive sharp bounds on the distribution of y^* under both corrupt and contaminated sampling models. Hotz, Mullins, and Sanders (1997), Kreider and Pepper (2007), Kreider and Hill (2008), and Kreider, Pepper, Gundersen and Jolliffe (2009) use this framework to derive bounds on the mean regression model when a regressor, x^* , is measured with error.

To illustrate how these bounds work, we focus on the important and relatively simple example of a binary outcome variable which may in some cases be misclassified.¹⁴ To simplify the exposition, regressors are assumed to be accurately measured and are left implicit; in any of the following one can condition the results on the observed regressors, x^* . Although our focus is on identification, the actual estimation strategy is very simple. Given a random sample from the joint distribution of (y, x^*) , the derived identification bounds can be consistently estimated by replacing population probabilities with their sample analogs.

In this context, we first present the corrupt and contaminated sampling bounds, and then

¹⁴As discussed in the previous section, when a variable with bounded support is imperfectly classified, it is widely recognized that the classical errors-in-variables model assumption of independence between measurement error and true variable cannot hold. Molinari (2008) presents an alternative and useful conceptualization of the data error problem for discrete outcome variables. This “direct misclassification” approach allows one to focus on assumptions related to classification error rates instead of restrictions on the mixing process.

apply these methods to a couple of important questions in criminal justice research. These examples are by no means exhaustive, but help give a taste for important considerations and implications when using error-ridden measurements in statistical analysis. Most notably, we observe that small amounts of measurement error can lead to substantial ambiguity about the true mean regression. Stronger assumptions on the unobserved error process can lead to sharper inferences, but may not be credible in many applications.

4.1 Mixture Model Bounds

Let y^* be an indicator function for self-reported drug use, and suppose one is interested in making inferences on the rate of illicit drug use: $P(y^* = 1) = E(y^*)$. Some unknown fraction of respondents, $P(y = 1, z = 0)$, inaccurately admit to using drugs (false positives) while another fraction, $P(y = 0, z = 0)$, inaccurately deny using drugs (false negatives). The relationship between the true and reported rates of drug use is as follows:

$$P(y^* = 1) = P(y = 1) + P(y = 0, z = 0) - P(y = 1, z = 0) \quad (15)$$

If the fraction of false negatives exactly offsets the fraction of false positives, then the reported rate of use equals the true rate of use: $P(y^* = 1) = P(y = 1)$. Unfortunately, the data alone only identify the fraction of the population that self-reports use: $P(y = 1)$. The sampling process cannot identify the fraction of false negatives or false positives.

A common starting point in this literature is to assume a known lower bound v on the fraction of cases that are drawn from the distribution of interest:¹⁵

$$P(z = 1) \geq v \quad (16)$$

A particular lower bound restriction may be informed by a validation study of a related population (e.g, Harrison, 1995) or the known fraction of responses that are imputed. Moreover, by varying the value of v , we can consider the wide range of views characterizing

¹⁵This type of restriction is used in the literatures on robust statistics (Huber, 1981) and data errors with binary regressors (see, e.g., Bollinger, 1996 and Frazis and Loewenstein, 2003).

the debate on inaccurate reporting. Those willing to assume fully accurate reporting can set $v = 1$, in which case the sampling process identifies the outcome probability. Those uncomfortable with placing any lower bound on the fraction of accurate responses can set $v = 0$, in which case the sampling process is uninformative. Middle ground positions are evaluated by setting v somewhere between 0 and 1.

Given the restriction that no more than some fraction, $1 - v$, of the population misreport, we know from Equation (15) that in the case of corrupt sampling, bounds are as follows:

$$\max\{P(y = 1) - (1 - v), 0\} \leq P(y^* = 1) \leq \min\{P(y = 1) + (1 - v), 1\} \quad (17)$$

Under contaminated sampling, where we are willing to assume that y^* and z are statistically independent, the bounds are different:

$$\max\{[P(y = 1) - (1 - v)]/v, 0\} \leq |Z = 1| \leq P(y^* = 1) \leq \min([P(y = 1)/v], 1). \quad (18)$$

These bounds are derived by Horowitz and Manski (1995: Corollary 1.2), and are referred to as the corrupt and contaminated sampling bounds, respectively.

Several features of these bounds are important. First, the bounds are sensitive to the upper bound misreporting rate, $1 - v$. Identification of $P(y^* = 1)$ deteriorates rapidly with the allowed fraction of mis-classifications, so small amounts of error can have large effects on inferences. Second, the contaminated sampling bounds are narrower than the corrupt sampling bounds. Whether the independence assumption is valid, however, depends on the application: it seems unlikely, for example, that the misreporting of illicit drug use is independent of actual drug use status or that the true crime rate is independent of whether an observation must be imputed. Finally, without additional assumptions, we cannot point identify the mean regression. Rather, all we can conclude is that the true outcome probability lies within some upper and lower bound.

4.2 Mixture Model Applications

To illustrate how the mixture model might be applied in practice, we consider two important questions in criminal justice research where non-classical measurement errors are likely to

confound inference: these are the problem of using data from the NSDUH to infer the rate of illicit drug use, and using data from the UCR to infer the aggravated assault rate. In the former case, the data are contaminated from response errors; some respondents do not provide accurate reports of drug use. In the latter case, the data are contaminated from imputation errors; about one-third of all reporting agencies, representing over ten-percent of the population, have imputed crime data.

4.2.1 Illustration 3: Illicit Drug Use

To illustrate the implications of data errors using the mixing model approach, consider using data from the NSDUH to draw inferences on the true rate of illicit drug use. While there is very little information on the degree of data errors in this survey, there are good reasons to believe that the errors are extensive and systematic. Respondents concerned about the legality of their behavior may falsely deny consuming illicit drugs, while the desire to fit into a deviant culture or otherwise be defiant may lead some respondents to falsely claim to consume illicit drugs (Pepper, 2001). Thus, ignoring these errors may be problematic and the classical errors-in-variables model is inappropriate.

Instead, Kreider and Pepper (forthcoming) consider using the mixing model to address the problem of drawing inferences on the rate of marijuana use in the presence of non-random reporting errors. The 2002 NSDUH reveals that 54% of 18-24 year-olds claimed to have consumed marijuana within their lifetime, with 30% reporting use during the last year (Office of Applied Studies, 2003). To draw inferences about true rates of illicit drug use in the United States, one must combine these self-reports with assumptions about the nature and extent of reporting errors. As noted above, Harrison (1995), who compares self-reported marijuana use to urinalysis test results among a sample of arrestees, finds a 22% misreporting rate for marijuana consumption. Arguably, the accurate reporting rate, z , in the general non-institutionalized population exceeds that obtained in the sample of arrestees studied by Harrison (1995). Arrestees have a relatively high incentive to misreport (Harrison, 1995; Pepper, 2001). Under this restriction alone, the corrupt sampling bounds reveal much

uncertainty about the true rates of drug use. For example, we only learn that between 32% (= 54 - 22) and 76% (= 54 + 22) of the young adult population has ever used marijuana. Importantly, this uncertainty reflects the identification problem caused by data errors; these bounds do not reflect the presence of additional uncertainty due to sampling variability.

Under the contaminated sampling assumption, the bounds narrow considerably. The bounds on lifetime marijuana use, for example, narrow from [32%, 76%] to [41%, 69%], a 36 percent reduction in bound width. When Kreider and Pepper impose the additional assumption that all draws from the alternative distribution, \tilde{y}^* , are in error (i.e., the response error model), the lifetime marijuana use rate is nearly point-identified, lying in the narrow range [54%, 57%]. These latter findings, however, rest on the implausible contaminated sampling assumption that drug use rates are identical among accurate and inaccurate reporters. More realistically, the rate of illicit drug use is higher among inaccurate reporters. Under this restriction, the lifetime rate of marijuana use is bounded to lie within [54%, 76%].

A useful practical feature of mixing model results is that we can assess the sensitivity of the bounds to variation in v . After all, Harrison's estimates might not accurately reflect misreporting rates in the general population. Figures 1 and 2 display bounds on the lifetime and past year marijuana use rates, respectively, under the corrupt sampling, contaminated sampling, and response error models considered by Kreider and Pepper (forthcoming). The vertical axis measures the outcome probability, $P(y = 1)$, and the horizontal axis represents the lower bound fraction of responses known to come from the distribution of interest, v .

[Figure 1 about here.]

[Figure 2 about here.]

The diagonal lines converging at $P(y = 1) = P(y^* = 1)$ when $v = 1$ depict the Horowitz and Manski corrupt sampling bounds. For any v , $P(y = 1)$ must lie within the vertical interval between these diagonal lines. The lower bound is uninformative for $v \leq 1 - P(y^* = 1)$, while the upper bound is uninformative for $v \leq P(y^* = 1)$. Thus, for lifetime use, the

bounds are uninformative unless we know that over 50% of responses are valid; for past year use, the lower bound is uninformative unless we know that over 30% of responses are valid. What is most striking about these corrupt sampling bounds is that even with fairly small degrees of reporting error, there is much ambiguity in the prevalence rate of marijuana use. If, for example, we know that only ten-percent of all reports may be misclassified, then the lifetime use rate can only be restricted to lie within a 20 point range: $[0.44, 0.64]$. Without richer information on the nature and degree of reporting errors in the NSDUH, the only way to draw tighter inferences is to impose additional assumptions.

It is tempting to address the data error problem with strong, but possibly flawed, modeling assumptions. The contaminated sampling models, for example, substantially narrow the corrupt sampling bounds (see Figures 1 and 2). As discussed above, however, these models are untenable in this setting. In this case, stronger assumptions do not resolve the ambiguity reflected in the corrupt sampling bounds—they simply replaced uncertainty over data errors with uncertainty over the model.

There may be other assumptions that can be credibly applied. Pepper (2001), for example, assumes the fraction of false negatives exceeds the fraction of false positives, in which case the reported lifetime rate of 0.54 serves as a lower bound for all v . Thus, if no more than 10% of respondents are misclassified, the lifetime prevalence rate is bounded to lie with $[0.54, 0.64]$. Kreider and Pepper (forthcoming) formalize the assumption that the rate of illicit drug use is higher among inaccurate reporters. While these assumption narrow the bounds, there remains much uncertainty about the rates of use and, as shown in Pepper (2001), identifying trends in use can be even more problematic.

4.2.2 Illustration 4: The Aggravated Assault Rate

Mixing models of measurement error are also a natural way to address imputation errors in UCR reported crime figures. While most police agencies provide UCR data to the FBI, Maltz and Targonski (2002) and Lynch and Jarvis (2008) find a non-trivial and non-random portion of UCR data are imputed. Consider, for example, the problem of inferring the 2005

aggravated assault rate in the United States from the rate of 0.0029 in the UCR (see Table 1).¹⁶ Without additional information, the corrupt sampling bounds reveal that violent crime rates lie between zero and the imputation rate. So, for example, if crime rates are imputed for 5% of the population, then the 2005 aggravated assault rate is only known to lie between [0, 5%]. Information identifying which records are imputed may narrow these bounds, but not by an appreciable amount.

By contrast, information on the direction of the imputation bias might be informative. Maltz and Targonski (2002) argue, for example, that nonresponse is likely to occur in periods where there is little reported crime. If so, this would lead to an upward imputation bias, so that the observed rate of 0.0029 serves as an upper bound on the 2005 aggravated assault rate. Unfortunately, this assumption seems unsubstantiated. Arguably, some police departments do not respond (or under-report) in periods where crime rates are inflated, in which case there would be a downward imputation bias (Mosher, Miethe, and Phillips, 2002).

Maltz and Targonski (2002) have argued that imputation errors may bias conclusions drawn from analyses conducted at fine geographic levels, such as counties, but will be less important for inferences about state and national-level crime rates. The data alone do not support this conclusion. Unless one is willing to make strong and seemingly unsubstantiated assumptions, even small amounts of imputation errors can lead to substantial uncertainties about true crime rates and trends.

5 Conclusion: The Law of Decreasing Credibility

Measurement errors continue to frustrate attempts to draw credible inferences from data used to track the extent and expression of crime in the United States. Lack of detailed information on the degree and nature of measurement errors in major national crime datasets, namely the UCR and NCVS, is especially troubling. In the absence of direct information on these

¹⁶In this discussion, we are concerned with drawing inferences on the true rate of aggravated assault reported to the police. Inferences regarding the overall rate of aggravated assault—known and unknown to the police—are complicated by the proxy variables problem discussed previously.

errors, inferences on crime rates and trends, and on the impact of policy on crime, are largely speculative. Although it might be—as some have suggested—that misreporting rates are stable over time and unrelated to policies of interest, this conjecture seems implausible and is unsupported by evidence. Either way, measurement errors are likely to be both substantial and systematic in survey data on crime and illicit behavior.

Though these problems do not imply that the data are completely uninformative, they do imply that researchers must choose between the unpleasant alternatives of either tolerating a certain degree of ambiguity in inference, or imposing strong assumptions about unobserved measurement errors. The problem, of course, is that weak assumptions may lead to indeterminate conclusions, whereas strong assumptions may be inaccurate and yield flawed conclusions (Pepper, 2001; Manski, 2007; Manski, Newman and Pepper, 2000). Manski (2007) refers to this fundamental trade-off as the *Law of Decreasing Credibility*: stronger assumptions yield sharper but less credible inferences.

This trade-off should not be easily dismissed. Imposing convenient assumptions does not resolve the measurement error problem, but simply exchanges uncertainty over unobserved errors with uncertainty over the accuracy of the model. Assumptions that data errors are exogenous or “classical,” for example, are in many applications untenable. As we have noted in this chapter, relaxing the central assumptions of the classical errors-in-variable model has substantive implications for the conclusions we might draw from the data. Inferences are highly sensitive to even small amounts of measurement and modeling errors.

There are practical solutions to this predicament. If stronger assumptions are not imposed, the way to resolve an indeterminate finding is to collect richer data. More detailed information on the nature of data error problems might supplement the existing data and help to suggest credible assumptions about error processes. Alternatively, efforts to increase the valid response rate may directly reduce the potential effects of these problems. Even with the best survey sampling methods, however, researchers must confront the fact that data on such sensitive topics as crime and victimization will always be subject to poorly behaved

measurement errors, and inferences drawn using these data will be impacted by such errors. Failure to seriously address data error problems can only lead to decreased credibility and potentially costly mistakes in drawing inferences relevant to crime policy.

References

- [1] Anglin, M.D., J.P. Caulkins, and Y. Hser. (1993). "Prevalence Estimation: Policy Needs, Current Status, and Future Potential," *Journal of Drug Issues*, 23(2): 345-360.
- [2] Azrael, D., P.J. Cook, and M. Miller. (2001). "State and Local Prevalence of Firearms Ownership: Measurement Structure and Trends," *National Bureau of Economic Research: Working Paper* 8570.
- [3] Bennett, T., K. Holloway, and D. Farrington. (2008). "The Statistical Association Between Drug Misuse and Crime: A Meta-Analysis," *Aggression and Violent Behavior*, 13: 107-118.
- [4] Black, D. (1970). "Production of Crime Rates," *American Sociological Review*, 35: 733-48.
- [5] Black, D.A., Berger, M.C. and F.A. Scott. (2000). "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 95(451): 739-748.
- [6] Blumstein and Rosenfeld. (2009). *Factors Contributing to U.S. Crime Trends, in Understanding Crime Trends: Workshop Report*, Arthur S. Goldberger and Richard Rosenfeld, eds., Committee on Understanding Crime Trends, Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.).
- [7] Bollinger, C. (1996). "Bounding Mean Regressions When a Binary Variable is Mismeasured," *Journal of Econometrics*, 73(2): 387-99.
- [8] Bound, J., C. Brown, and N. Mathiowetz (2001). "Measurement Error in Survey Data," In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, 5, Ch. 59: 3705-3843.
- [9] Chaiken, J.M. and M.R. Chaiken. (1990). "Drugs and Predatory Crime," *Crime and Justice: Drugs and Crime*, 13: 203-239.
- [10] Dominitz, J. and R. Sherman (2004). "Sharp Bounds Under Contaminated or Corrupted Sampling With Verification, with an Application to Environmental Pollutant Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 9(3): 319-338.

- [11] Frazis, H. and M. Loewenstein (2003). "Estimating Linear Regressions with Mismeasured, Possibly Endogenous, Binary Explanatory Variables," *Journal of Econometrics*, 117: 151-178.
- [12] Frisch, R., 1934, *Statistical confluence analysis by means of complete regression systems* (University Institute for Economics, Oslo).
- [13] Griliches, Z. and J.A. Hausman. (1985). "Errors in Variables in Panel Data: A Note with an Example," *Journal of Econometrics*, 31(1): 93-118.
- [14] Harrison, L. D. (1995). "The Validity of Self-Reported Data on Drug Use," *Journal of Drug Issues*, 25(1): 91-111.
- [15] Harrison, L. and A. Hughes (1997). "Introduction - The validity of self-reported drug use: Improving the accuracy of survey estimates." In Harrison and Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates*. NIDA Research Monograph, 167, 1-16. Rockville, MD: US Department of Health and Human Services.
- [16] Horowitz, J. and C. Manski (1995). "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63(2): 281-02.
- [17] Huber, P. (1981). *Robust Statistics*. New York: Wiley.
- [18] Hotz, J., C. Mullins, and S. Sanders (1997). "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," *Review of Economic Studies*, 64(4), 575-603.
- [19] Johnston, L.D., O'Malley, P.M., and J.G. Bachman. (1998). *National survey results on drug use from the Monitoring the Future study, 1975-1997, Volume I: Secondary school students*. NIH Publication No. 98-4345. Rockville, MD: National Institute on Drug Abuse.
- [20] Kleck, G. and M. Gertz. (1995). "Armed resistance to crime: The prevalence and nature of self-defense with a gun," *Journal of Criminal Law and Criminology*, 86: 150-187.
- [21] Klepper S. and E.E. Leamer. (1984). "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52(1): 163-183.
- [22] Koss, M. (1993). "Detecting the scope of rape: A review of prevalence research methods," *Journal of Interpersonal Violence*, 8: 198-222.
- [23] _____. (1996). "The measurement of rape victimization in crime surveys," *Criminal Justice and Behavior*, 23: 55-69.

- [24] Kreider, B. and S. Hill. (forthcoming). “Partially Identifying Treatment Effects with an Application to Covering the Uninsured,” *Journal of Human Resources*.
- [25] Kreider, B. and J. Pepper (2007). “Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors,” *Journal of the American Statistical Association*, 102 (478): 432-441.
- [26] _____ and _____. (2008). “Inferring Disability Status from Corrupt Data,” *Journal of Applied Econometrics*, 23(3), 329-49.
- [27] _____ and _____. (forthcoming). “Identification of Expected Outcomes in a Data Error Mixing Model with Multiplicative Mean Independence,” *Journal of Business and Economic Statistics*.
- [28] Kreider, B., J. Pepper, C. Gundersen, and D. Jolliffe. (2009). “Identifying the Effects of Food Stamps on Children’s Health Outcomes When Participation is Endogenous and Misreported,” Working Paper.
- [29] Lambert, D. and L. Tierney. (1997). “Nonparametric Maximum Likelihood Estimation from Samples with Irrelevant Data and Verification Bias,” *Journal of the American Statistical Association*, 92: 937-944.
- [30] Lynch, J.P. and L.A. Addington (Eds.). (2007). *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, Cambridge: Cambridge University Press.
- [31] Lynch, J. and J. Jarvis. (2008). “Missing Data and Imputation in the Uniform Crime Reports and the Effects on National Estimates.” *Journal of Contemporary Criminal Justice*, Vol. 24(1): 69-85. DOI: 10.1177/1043986207313028.
- [32] Maltz, M. (1999). “Bridging Gaps in Police Crime Data: A Discussion Paper from the BJS Fellows Program,” *Bureau of Justice Statistics*, Government Printing Office, Washington DC.
- [33] Maltz, M.D. and J. Targonski. (2002). “A note on the use of county-level UCR data.” *Journal of Quantitative Criminology*. 18:297-318.
- [34] Manski, C.F. (2007). *Identification for Prediction and Decisions*. Harvard University Press, Cambridge, MA.
- [35] Manski, C.F., J. Newman and J.V. Pepper. (2002). “Using Performance Standards to Evaluate Social Programs with Incomplete Outcome Data: General Issues and Application to a Higher Education Block Grant Program,” *Evaluation Review*, 26(4), 355-381.

- [36] McDowall, D. and C. Loftin. (2007). "What is Convergence, and What Do We Know About It?" In J. Lynch and L.A. Addington, *Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR*, Ch. 4: 93-124.
- [37] McDowall, D., C. Loftin, and B. Wierseman. (1998). "Estimates of the Frequency of Firearm Self-Defense from the Redesigned National Crime Victimization Survey." Violence Research Group Discussion Paper 20.
- [38] Molinari, F. (2008). "Partial Identification of Probability Distributions with Misclassified Data," *Journal of Econometrics*, 144(1): 81-117.
- [39] Mosher, C.J., T.D. Miethe, and D.M. Phillips. (2002). *The Mismeasure of Crime*. Sage Publications, Thousand Oaks, CA.
- [40] Mullin, C.H. (2005). "Identification and Estimation with Contaminated Data: When do covariate Data Sharpen Inference?" *Journal of Econometrics*, 130: 253-272.
- [41] National Research Council. (2001). *Informing America's Policy on Illegal Drugs: What We Don't Know Keeps Hurting Us*. Committee on Data and Research for Policy on Illegal Drugs. Charles F.Manski, John V.Pepper, and Carol V.Petrie, editors. Committee on Law and Justice and Committee on National Statistics. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- [42] _____. (2003). *Measurement Problems in Criminal Justice Research: Workshop Summary*. J.V. Pepper and C.V. Petrie. Committee on Law and Justice and Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- [43] _____. (2005). *Firearms and Violence: A Critical Review*. Committee to Improve Research Information and Data on Firearms. Charles F. Wellford, John V. Pepper, and Carol V. Petire, editors. Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.
- [44] _____. (2008). *Surveying Victims: Options for Conducting the National Crime Victimization Survey*. Panel to Review the Programs of the Bureau of Justice Statistics. Robert M. Groves and Daniel L. Cork, eds. Committee on National Statistics and Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- [45] Office of Applied Studies. (2003). *Results from the 2002 National Survey on Drug Use and Health: Summary of National Finding*, (DHHS Publication No. SMA 03-3836, Series H-22). Rockville, MD: Substance Abuse and Mental Health Services Administration.

- [46] Pepper, J.V. (2001). "How Do Response Problems Affect Survey Measurement of Trends in Drug Use?" in C. F. Manski, J. V. Pepper, and C. Petrie (Eds.), *Informing America's Policy on Illegal Drugs: What We Don't Know Keeps Hurting Us*, National Academy Press, Washington, D.C., 321-48.
- [47] Rand, M.R. and C.M. Rennison. (2002). "True Crime Stories? Accounting for Differences in Our National Crime Indicators," *Chance*, 15(1): 47-51.
- [48] Tourangeau, R. and M.E. McNeeley. (2003). "Measuring Crime and Crime Victimization: Methodological Issues," in *Measurement Problems in Criminal Justice Research: Workshop Summary*. J.V. Pepper and C.V. Petrie. Committee on Law and Justice and Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- [49] U.S. Department of Justice. (2004). "The Nation's Two Crime Measures," NCJ 122705, <http://www.ojp.usdoj.gov/bjs/abstract/ntmc.htm>.
- [50] _____. (2006). *National Crime Victimization Survey: Criminal Victimization, 2005*. Bureau of Justice Statistics Bulletin. <http://www.ojp.usdoj.gov/bjs/pub/pdf/cv05.pdf>
- [51] _____. (2008). *Crime in the United States, 2007*. Federal Bureau of Investigation, Washington, DC. (table 1). http://www.fbi.gov/ucr/cius2007/data/table_01.html
- [52] Wansbeek, T. and E. Meijer. (2000). *Measurement Error and Latent Variables in Econometrics*. Amsterdam: Elsevier.
- [53] Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Table 1: UCR and NCVS Annual Crime Rates (per 1,000) in the United States: 1990, 2000 and 2005*

Crime Survey	Rape		Robbery		Assault		Property Crime	
	UCR	NCVS	UCR	NCVS	UCR	NCVS	UCR	NCVS
1990	0.4	1.7	2.6	5.7	4.2	9.8	50.7	348.9
2000	0.3	0.6	1.5	3.2	3.2	5.7	36.2	178.1
2005	0.3	0.5	1.4	2.6	2.9	4.3	34.3	154
% Change**	-22.6	-70.6	-45.1	-54.4	-31.2	-56.1	-32.4	-55.9

* UCR estimates come from the U.S. Department of Justice (2008), and NCVS estimates from the U.S. Department of Justice (2006).

** Percentage Change from 1990-2005.

Figure 1: Bounds on Lifetime Marijuana Use Given Response Error [$P(y^* = 1) = 0.54$]

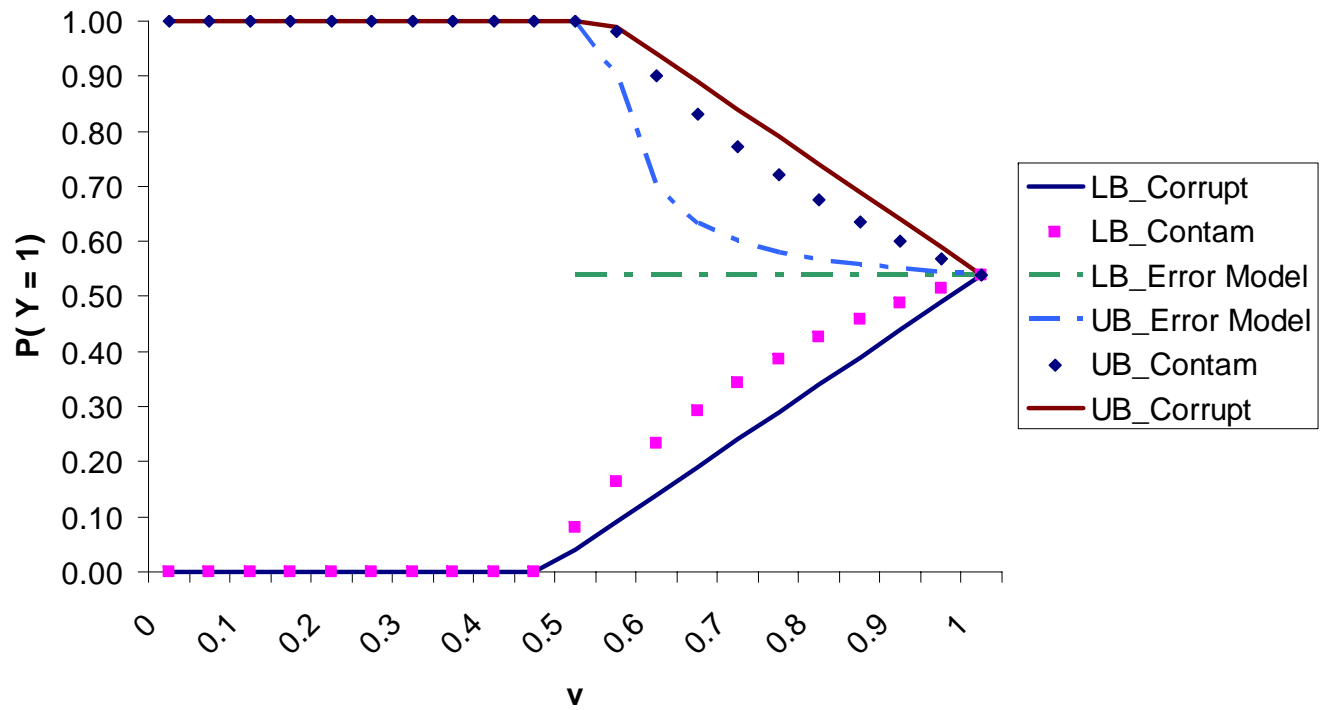


Figure 2: Bounds on Past Year Marijuana Use Given Response Errors [$P(y^* = 1) = 0.30$]

