Do you believe that every time a prisoner is executed in the United States, eight future murders are deterred? Do you believe that a 1% increase in the number of citizens licensed to carry concealed weapons causes a 3.3% decrease in the state's murder rate? Do you believe that 10 to 20% of the decline in crime in the 1990s was caused by an increase in abortions in the 1970s? Or that the murder rate would have increased by 250% since 1974 if the United States had not built so many new prisons?

If you were misled by any of these studies, you may have fallen for a pernicious form of junk science: the use of mathematical models with no demonstrated predictive capability to draw policy conclusions. These studies are superficially impressive. Written by reputable social scientists from prestigious institutions, they often appear in peer reviewed scientific journals. Filled with complex statistical calculations, they give precise numerical "facts" that can be used as debaters' points in policy arguments. But these "facts" are will o' the wisps. Before the ink is dry on one study, another appears with completely different "facts." Despite their scientific appearance, these models do not meet the fundamental criterion for a useful mathematical model: the ability to make predictions that are better than random chance.

Although economists are the leading practitioners of this arcane art, sociologists, criminologists and other social scientists have versions of it as well. It is known by various names, including "econometric modeling," "structural equation modeling," and "path analysis." All of these are ways of using the correlations between variables to make causal inferences. The problem with this, as anyone who has had a course in statistics knows, is that correlation is not causation. Correlations between two variables are often "spurious" because they are caused by some third variable. Econometric modelers try to overcome this problem by including all the relevant variables in their analyses, using a statistical technique called "multiple regression." If one had perfect measures of all the causal variables, this would work. But the data are never good enough. Repeated efforts to use multiple regression to achieve definitive answers to public policy...
questions have failed. But many social scientists are reluctant to admit failure. They have devoted years to learning and teaching regression modeling, and they continue to use regression to make causal arguments that are not justified by their data. I call these arguments the myths of multiple regression, and I would like to use four studies of murder rates as examples.

**Myth One: More Guns, Less Crime.**

John Lott, an economist at Yale University, used an econometric model to argue that "allowing citizens to carry concealed weapons deters violent crimes, without increasing accidental deaths." Lott's analysis involved "shall issue" laws that require local authorities to issue a concealed weapons permit to any law-abiding citizen who applies for one. Lott estimated that each one percent increase in gun ownership in a population causes a 3.3% decrease in homicide rates. Lott and his co-author, David Mustard posted the first version of their study on the Internet in 1997 and tens of thousands of people downloaded it. It was the subject of policy forums, newspaper columns, and often quite sophisticated debates on the World Wide Web. In a book with the catchy title More Guns, Less Crime, Lott taunted his critics, accusing them of putting ideology ahead of science.

Lott's work is an example of statistical one-upmanship. He has more data and a more complex analysis than anyone else studying the topic. He demands that anyone who wants to challenge his arguments become immersed in a very complex statistical debate, based on computations so difficult that they cannot be done with ordinary desktop computers. He challenges anyone who disagrees with him to download his data set and redo his calculations, but most social scientists do not think it worth their while to replicate studies using methods that have repeatedly failed. Most gun control researchers simply brushed off Lott and Mustard's claims and went on with their work. Two highly respected criminal justice researchers, Frank Zimring and Gordon Hawkins (1997) wrote an article explaining that:

just as Messrs. Lott and Mustard can, with one model of the determinants of homicide, produce statistical residuals suggesting that 'shall issue' laws reduce homicide, we expect that a determined econometrician can produce a treatment of the same historical periods with different models and opposite effects. Econometric modeling is a double-edged sword in its capacity to facilitate statistical findings to warm the hearts of true believers of any stripe.

Zimring and Hawkins were right. Within a year, two determined econometricians, Dan Black and Daniel Nagin (1998) published a study showing that if they changed the statistical model a little bit, or applied it to different segments of the data, Lott and Mustard's findings disappeared. Black and Nagin found that when Florida was removed from the sample there was "no detectable impact of the right-to-carry laws on the rate of murder and rape." They concluded that "inference based on the Lott and Mustard model is inappropriate, and their results cannot be used responsibly to formulate public policy."

John Lott, however, disputed their analysis and continued to promote his own. Lott had collected data for each of America's counties for each year from 1977 to 1992. The problem with this is that America's counties vary tremendously in size and social characteristics. A few large ones, containing major cities, account for a very large percentage of the murders in the United States. As it happens, none of these very large counties have "shall issue" gun control laws. This means that Lott's massive data set was simply unsuitable for his task. He had no variation in his key causal variable -- "shall issue" laws -- in the places where most murders occurred.

He did not mention this limitation in his book or articles. When I discovered the lack of "shall issue" laws in the major cities in my own examination of his data, I asked him about it. He shrugged it off, saying that he had "controlled" for population size in his analysis. But introducing a statistical control in the mathematical analysis did not make up for the fact that he simply had no data for the major cities where the homicide problem was most acute.

It took me some time to find this problem in his data, since I was not familiar with the gun control issue. But Zimring and Hawkins zeroed in on it immediately because they knew that "shall issue" laws were instituted in states where the National Rifle Association was powerful, largely in the South, the West and in rural regions. These were states that already had few restrictions on guns. They observed that this legislative history frustrates "our capacity to compare trends in 'shall issue' states with trends in other states. Because the states that changed legislation are different in location and constitution from the states that did not, comparisons across legislative categories will always risk confusing demographic and regional influences with the behavioral impact of different legal regimes." Zimring and Hawkins further observed that:

Lott and Mustard are, of course, aware of this problem. Their solution, a standard econometric technique, is to build a statistical model that will control for all the differences between Idaho and New York City that influence homicide and crime rates, other than the "shall issue" laws. If one can "specify" the major influences on homicide, rape, burglary, and auto theft in our model, then we can eliminate the influence of these factors on the different trends. Lott and Mustard build models that estimate the effects of demographic data, economic data, and criminal punishment on various offenses. These models are the ultimate in statistical home cooking in that they are created for this data set by these authors and only tested on the data that will be used in the evaluation of the right-to-carry impacts.

Lott and Mustard were comparing trends in Idaho and West Virginia and Mississippi with trends in Washington, D.C. and New York City. What actually happened was that there was an explosion of crack-related homicides in major eastern cities in the 1980s and early 1990s. Lott's whole argument came down to a claim that the largely rural and western "shall issue" states were spared the crack-related homicide epidemic because of their "shall issue" laws. This would never have been taken seriously if it had not been obscured by a maze of equations.

**Myth Two: Imprisoning More People Cuts Crime**
The Lott and Mustard case was exceptional only in the amount of public attention it received. It is quite common, even typical, for rival studies to be published using econometric methods to reach opposite conclusions about the same issue. Often there is nothing demonstrably wrong with either of the analyses. They simply use slightly different data sets or different techniques to achieve different results. It seems as if regression modelers can achieve any result they want without violating the rules of regression analysis in any way.

In one exceptionally frank statement of frustration with this state of affairs, two highly respected criminologists, Thomas Marvell and Carlisle Moody (1997: 221), reported on the reception of a study they did of the effect of imprisonment on homicide rates. They reported that they:

widely circulated [their] findings, along with the data used, to colleagues who specialize in quantitative analysis. The most frequent response is that they refuse to believe the results no matter how good the statistical analysis. Behind that contention is the notion, often discussed informally but seldom published, that social scientists can obtain any result desired by manipulating the procedures used. In fact, the wide variety of estimates concerning the impact of prison populations is taken as good evidence of the malleability of research. The implication, even among many who regularly publish quantitative studies, is that no matter how thorough the analysis, results are not credible unless they conform with prior expectations. A research discipline cannot succeed in such a framework.

To their great merit, Marvell and Moody frankly acknowledged the problems with multiple regression, and made some suggestions for improvement. Unfortunately, some econometricians become so immersed in their models that they lose track of how arbitrary they are. They come to believe that their models are more real, more valid, than the messy, recalcitrant, "uncontrolled" reality they purport to explain.

**Myth Three: Executing People Cuts Crime**

In 1975 The American Economic Review published an article by a leading economist, Isaac Ehrlich of the University of Michigan, who estimated that each execution deterred eight homicides. Before Ehrlich, the best known specialist on the effectiveness of capital punishment was Thorsten Sellen, who had used a much simpler method of analysis. Sellen prepared graphs comparing trends in different states. He found little or no difference between states with or without the death penalty, so he concluded that the death penalty made no difference. Ehrlich, in an act of statistical one-upmanship, claimed that his analysis was more valid because it controlled for all the factors that influence homicide rates.

Even before it was published, Ehrlich's work was cited by the Solicitor General of the United States in an amicus curiae brief filed with the United States Supreme Court in defense of the death penalty. Fortunately, the Court decided not to rely upon Ehrlich's evidence because it had not been confirmed by other researchers. This was wise, because within a year or two other researchers published equally sophisticated econometric analyses showing that the death penalty had no deterrent effect.

The controversy over Ehrlich's work was so important that the National Research Council convened a blue ribbon panel of experts to review it. After a very thorough review, the panel decided that the problem was not just with Ehrlich's model, but with the idea of using of econometric methods to resolve controversies over criminal justice policies. They (Manski, 1978: 422) concluded that:

because the data likely to be available for such analysis have limitations and because criminal behavior can be so complex, the emergence of a definitive behavioral study lying to rest all controversy about the behavioral effects of deterrence policies should not be expected.

Most experts now believe that Sellen was right, that capital punishment has no demonstrable effect on murder rates. But Ehrlich has not been persuaded. He is now a lonely true believer in the validity of his model. In a recent interview (Bonner and Fessendren, 2000) he insisted "if variations like unemployment, income inequality, likelihood of apprehension and willingness to use the death penalty are accounted for, the death penalty shows a significant deterring effect."
Myth Four: Legalized Abortion Caused the Crime Drop in the 1990s.

In 1999, John Donohue and Steven Levitt released a study with a novel explanation of the sharp decline in murder rates in the 1990s. They argued that the legalization of abortion by the U.S. Supreme Court in 1973 caused a decrease in the birth of unwanted children, a disproportionate number of whom would have grown up to be criminals. The problem with this argument is that the legalization of abortion was a one-time historical event and one-time events do not provide enough data for a valid regression analysis. It is true that abortion was legalized earlier in some states than others, and Donohue and Levitt make use of this fact. But all these states were going through the same historical processes, and many other things were happening in the same historical period that effected murder rates. A valid regression analysis would have to capture all of these things, and test them under a wide range of variation. The existing data do not permit that, so the results of a regression analysis will vary depending on which data are selected for analysis.

In this case, Donohue and Levitt chose to focus on change over a twelve year time span, ignoring fluctuations within those years. By doing this, as James Fox (2000: 303) pointed out, "they missed most of the shifts in crime during this period - the upward trend during the late 1980s crack era and the downward correction in the post-crack years. This is something like studying the effects of moon phases on ocean tides but only recording data for periods of low tide."

When I was writing this article, I included a sentence stating "soon another regression analyst will probably reanalyze the same data and reach different conclusions." A few days later, my wife handed me a newspaper story about just such a study. The author was none other than John Lott of Yale, together with John Whitley of the University of Adelaide. They crunched the same numbers and concluded that "legalizing abortion increased murder rates by around about 0.5 to 7 percent" (Lott and Whitely, 2001).

Why such markedly different results? Each set of authors simply selected a different way to model an inadequate body of data. Econometrics cannot make a valid general law out of the historical fact that abortion was legalized in the 1970s and crime went down in the 1990s. We would need at least a few dozen such historical experiences for a valid statistical test.

Conclusions.

The acid test in statistical modeling is prediction. Prediction does not have to be perfect. If a model can predict significantly better than random guessing, it is useful. For example, if a model could predict stock prices even slightly better than random guessing, it would make its owners very wealthy. So a great deal of effort has gone into testing and evaluating models of stock prices. Unfortunately, researchers who use econometric techniques to evaluate social policies very seldom subject their models to predictive tests. Their excuse is that it takes too long for the outcomes to be known. You don’t get new data on poverty, abortion or homicide every few minutes as you do with stock prices. But researchers can do predictive testing in other ways. They can develop a model using data from one jurisdiction or time period, then use it to predict data from other times or places. But most researchers simply do not do this, or if they do the models fail and the results are never published.

The journals that publish econometric studies of public policy issues often do not require predictive testing, which shows that the editors and reviewers have low expectations for their fields. So researchers take data for a fixed period of time and keep fine tuning and adjusting their model it until they can "explain" trends that have already happened. There are always a number of ways to do this, and with modern computers it is not terribly hard to keep trying until you find something that fits. At that point, the researcher stops, writes up the findings, and sends the paper off for publication. Later, another researcher may adjust the model to obtain a different result. This fills the pages of scholarly journals, and everybody pretends not to notice that little or no progress is being made. But we are no closer to having a valid econometric model of murder rates today than we were when Isaac Ehrlich published the first model in 1975.
In 1991, David Freedman, a distinguished sociologist at the University of California at Berkeley and the author of textbooks on quantitative research methods, shook the foundations of regression modeling when he frankly stated "I do not think that regression can carry much of the burden in a causal argument. Nor do regression equations, by themselves, give much help in controlling for confounding variables" (Freedman, 1991: 292). Freedman's article provoked a number of strong reactions. Richard Berk (1991: 315) observed that Freedman's argument "will be very difficult for most quantitative sociologists to accept. It goes to the heart of their empirical enterprise and in so doing, puts entire professional careers in jeopardy."

Faced with critics who want some proof that they can predict trends, regression modelers often fall back on statistical one-upmanship. They make arguments so complex that only other highly trained regression analysts can understand, let alone refute, them. Often this technique works. Potential critics simply give up in frustration. The Philadelphia Inquirer's David Boldt (1999), after hearing John Lott speak on concealed weapons and homicide rates, and checking with other experts, lamented that "trying to sort out the academic arguments is almost a fool's errand. You can drown in disputes over t-statistics, dummy variables and 'Poisson' vs. 'least squares' data analysis methods."

Boldt was correct to suspect that he was being lured into a fool's mission. There are, in fact, no important findings in sociology or criminology that cannot be communicated to journalists and policy makers who lack graduate degrees in econometrics. It is time to admit that the emperor has no clothes. When presented with an econometric model, consumers should insist on evidence that it can predict trends in data other than the data used to create it. Models that fail this test are junk science, no matter how complex the analysis.

REFERENCES


Myth 3: Incidents of mass murder are increasing. When a mass murder occurs, it receives instant and pervasive news coverage. Unfortunately, we are prone to overestimate the frequency of an event by its prominence in our minds, and mass murder is no exception. This is a very rare phenomenon and is neither increasing nor decreasing in the US. The majority of adult mass murderers are psychotic, meaning they have broken with consensual reality and perceive the world in an idiosyncratic and often paranoid way. Yet they may research the internet for weapons, practice video games to sharpen their marksmanship, purchase weapons and ammunition, conduct surveillance of the target, and carry out their mass murder, all from within a delusion. "Myths of murder and multiple regression". According to the Trace Report, about 40 percent of illegally used or possessed firearms recovered in Chicago from 2013 to 2016 came from dealers in Illinois. The remaining 60 percent came from states with less regulation over firearms. The stats you were using were total murder and manslaughter rates, which could also involve other weapons like knives. On the technology side, police in 12 districts are using ShotSpotter, a sophisticated system of electronic monitoring to detect gunshots and alert officers to the location within 30 to 45 seconds.