Econometric Modeling as Junk Science

http://www.crab.rutgers.edu/~goertzel/mythsofmurder.htm

by Ted Goertzel, Rutgers University, Camden NJ 08102
goertzel@camden.rutgers.edu
voice: 609 953-1670 fax: 413 793-2597

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 percentage of a state's citizens carrying 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 build so many new prisons? Did you believe predictions that the welfare reform of the 1990s would force 1,100,000 children into poverty?

If you were misled by any of these studies, you may have fallen for a pernicious form of junk science: the use of mathematical modeling to evaluate the impact of social policies. These studies are superficially impressive. Produced by reputable social scientists from prestigious institutions, they are often published in peer reviewed scientific journals. They are filled with statistical calculations too complex for anyone but another specialist to untangle. They give precise numerical "facts" that are often quoted in policy debates. But these "facts" turn out to be will o' the wisps. Often before the ink is dry on one apparently definitive study, another appears with equally precise and imposing, but completely different, "facts." Despite their numerical precision, these "facts" have no more validity than the visions of soothsayers.

These predictions are based on a statistical technique called multiple regression that uses correlational analysis to make causal arguments. 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,” “path analysis” and simply “multivariate analysis.” All of these are all ways of using correlational data to make causal arguments.

The problem with this, as anyone who has studied statistics knows, is that correlation is not causation. A correlation between two variables may be “spurious” if it is caused by some third variable. Multiple regression researchers try to overcome the spuriousness problem by including all the variables in analysis. The data available for this purpose simply is not up to this task, however, and the studies have consistently failed. But many social scientists have devoted years to learning and teaching regression modeling. They continue to use regression to make causal arguments that are not justified by their data, but that get repeated over and over in policy arguments. I call these arguments the myths of multiple regression.

Five Myths of Multiple Regression

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 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 released 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. The debate followed predictable ideological lines, with one prominent critic denouncing the study as methodologically flawed before she had even received a copy. 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 argument, based on a data set that is so large that it cannot even be manipulated with the desktop computers most social scientists use. He is glad to share his data with any researcher who wants to use it, but most social scientists have tired of this game. How much time should researchers spend replicating and criticizing 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: 57) 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."

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 set of data. 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 esti-*

mates 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. This, however, is more the exception than the rule with econometricians, who often become so immersed in their models that they lose track of how arbitrary they are. Many of them 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 use 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.

Ehrlich was not persuaded by these critics, and found flaws in their work. He remains a lonely true believer in the validity of his model. In a recent interview (Bonner and Fessendren, 2000) he insisted that "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 is that the legalization of abortion was a one-time historical event and there are too little data for a valid regression analysis. The results are likely to vary depending on how data are selected for analysis. In this case, as James Fox (2000: 303) pointed out: "by employing a single statistic summarizing change over this twelve-year span, [Donohue and Levitt] miss 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 analyze 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 meaningful statistical test.

Myth Five: Welfare Reform Will Throw a Million Children into Poverty.

On August 1, 1996, as the United States Senate considered an epochal change in welfare policies, the Urban Institute issued a widely publicized report claiming to demonstrate that: *the proposed welfare reform changes would increase poverty and reduce incomes of families in the lowest income group... We estimate that 2.6 million more persons would fall below the poverty line as a result, including 1.1 million children*. (Urban Institute, 1996, p. 1)

Welfare advocates rallied around this prediction, but policy makers were not persuaded. Senators who supported the reform simply did not believe that social scientists could make valid predictions of that sort. And they were right. The Urban Institute could not even predict the direction of change, let alone its magnitude. Child poverty went down, not up, after the welfare reform.

The Urban Institute's model was much more complex than the other models we have examined in this paper, but the added complexity seems only to have compounded the problem. Using sophisticated "microsimulation" techniques, they took correlations that existed in the past, fed them into complex equations, then treated these equations as general laws. All their mathematics was based on the assumption that nothing fundamental would change, in which case, of course, welfare reform would fail. All the model did was produce numbers to illustrate their arguments and make them appear scientific. But the point of the reform was to change things, and it did.

Why Regression Fails

Although they seem complicated, regression models are actually simplifications of the real world. To simplify the mathematics, regression uses linear equations. This means it assumes that if you plot the relationship between any two variables on a graph, the trend will look like a straight line. Regression models also assume that variables are distributed according to a classic bell-shaped normal curve. And it assumes analysts know which variables are causes and which are effects. Of course, regression analysts know that the real world does not fit their assumptions, and they make various adjustments to the data to compensate. But the adjustments create other problems. The only valid way to test a model after all these adjustments is to show that it works to predict future trends. Regression models that have not been demonstrated to work with fresh data, other than the data used to create them, are junk science.

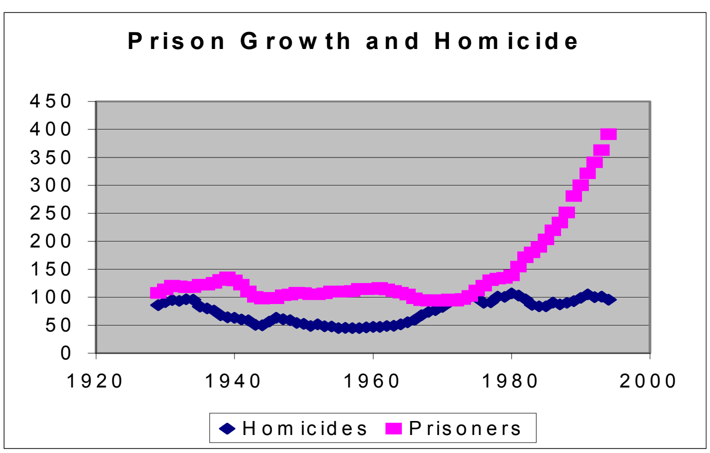
Why Regression Fails: Linear Models of a Nonlinear World.

When faced with the nonlinearity of the real world, the first instinct of the regression modeler is to standardize and control the data. In doing this, they minimize or eliminate the most interesting and important historical events. They end up analyzing a standardized and idealized world that bears little relationship to reality. For example, consider the trends in prison growth and homicide that Marvell and Moody (1997) sought to explain. Their paper begins with a graph showing trends in prisoners per 100,000 people and homicides per 1,000,000 people in the United States. This very interesting and useful graph is reproduced below from their data.

The interesting things about the trends the graph portrays are the turning points, the places where the trends diverge from linearity. Homicide rates increased sharply from the mid 1960s to the early 1970s, then leveled off. The number of prisoners shot up markedly beginning in the 1970s, as the United States built more prisons in response to the increasing crime rate. The homicide rate leveled off in the 1980s and remained stable thereafter.

Instead of trying to explain these important turning points, Marvell and Moody used multiple regression techniques to "control" for it. They introduced controls for every measurable variable they could think of, including (Marvell and Moody, 1997: 209) "age structure, economic factors, public relief, race, and variables marking World War II and the crack epidemic."

All these controls purged the striking historical changes from their data. This led them to the conclusion that a 10% increase in prison populations leads to roughly 13% fewer homicides. But a simple inspection of their graph shows that the promised 13% decline in the homicide rate for each 10% increase in imprisonment since 1975 simply did not occur.



Marvell and Moody were troubled that the expected reduction did not take place, but it was not enough to cause them to abandon their econometric methods. They, after all, were not discussing the real world but a world simplified and purified by a long series of mathematical adjustments. Confronted with the historical facts, they argued that, had imprisonment not increased, homicide would have gotten a lot worse. They went on, however, to observe that this would never really have happened because the government would have taken other actions to prevent it.

But what is the value of an analysis that leads to implications that the authors realize could never actually take place? How valid can the theory underling the multiple regression analysis be if it leaves out key variables, such as political constraints, simply because they cannot be quantitatively measured? How much do the results depend on arbitrary decisions about which control variables to introduce and how to measure them? Marvell and Moody were left with statistics that purported to tell us what might have happened if nothing that actually happened had happened.

Why Regression Fails: The World is Not a Bell-Shaped Curve.

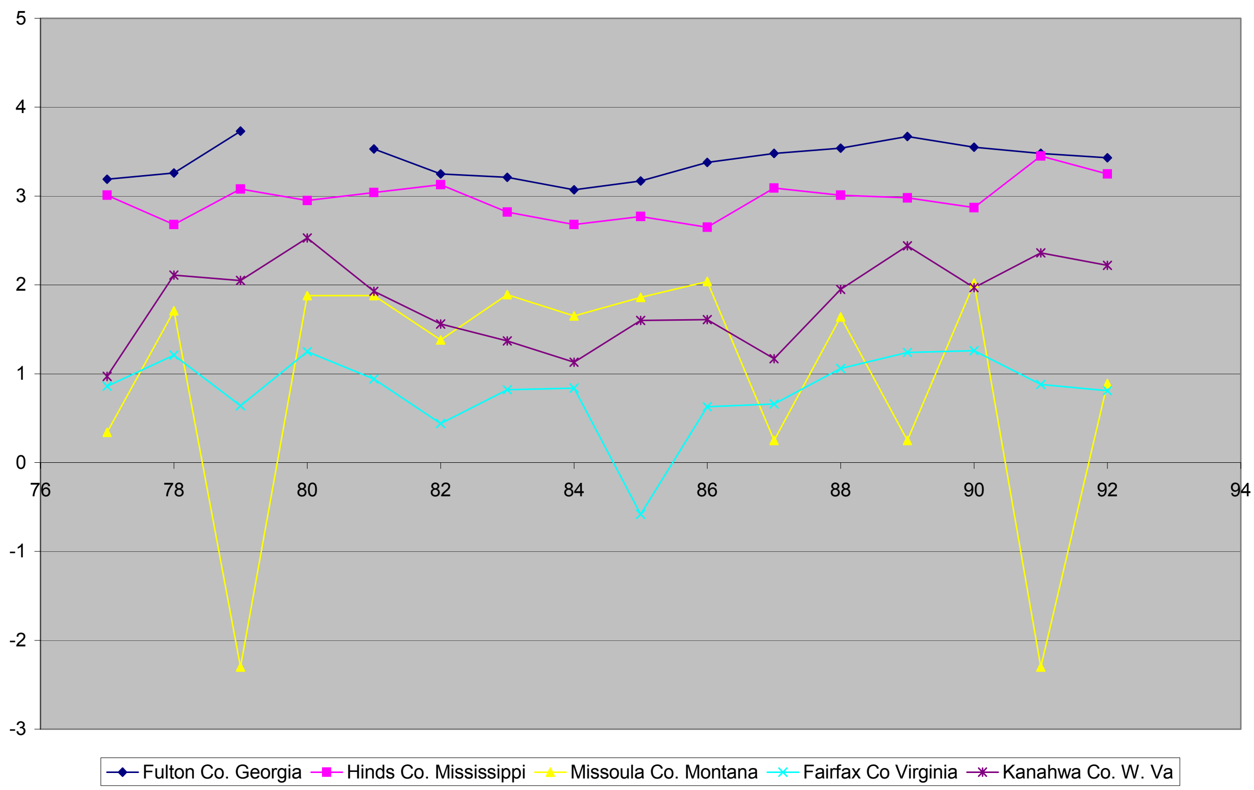
In addition to linearity, multiple regression assumes that each variable is "normally" distributed about all the others in a classic bell-curve pattern. This means that most cases should be clustered around the average within each category, with few at the extremes. Often the data violates this assumption in major ways that lead to completely erroneous results. A good example is John Lott's data on gun control.

Lott collected a massive data set that he generously made available to other researchers. Unfortunately, he did not begin by graphing his data, perhaps because he had so much of it. But it is always a good idea to begin an analysis with graphs so as to see the trends before they are obscured by all the statistical adjustments. If one cannot graph everything, it is still worthwhile to graph some representative cases. So, using Lott's data, I plotted trends in murder rates for a number of counties where "shall issue" laws had gone into effect during the period covered by his study. Before "shall issue" laws were passed, local officials had discretion in granting permits to carry concealed weapons. After they were passed, they had to issue a permit to any law abiding adult who wanted one. If Lott's hypothesis were correct, we would expect to see the murder rate go down once the laws were passed.

The following graph shows trends in murder rates for the largest counties in several states that adopted "shall issue" laws between 1977 and 1992. The date at which the laws went into effect varied from state to state. Before reading further, the reader may find it interesting to examine the graph and try to infer when the law took effect in each county.

Examining the graph, we see that the pattern in Missoula County, Montana, appears to be quite erratic, with very sharp declines in the murder rate in 1979 and 1991. This, however, is not a real phenomenon, but the result of one of the adjustments Lott made to compensate for the non-normality of his data. Instead of using the actual numbers, he converted his numbers to natural logarithms. This is a common practice, since natural logarithms often fit the assumptions of multiple regression better than the actual data. The number in John Lott's data file for Missoula County in those years is -2.30. This is odd, since a County's murder rate cannot go below zero, unless previously murdered people are brought back to life. No such luck, however. To get the actual murder rate in each county, one has to invert the logarithms in Lott's data set with the formula ***true rate* = e** $^{\textit{logarithmic rate}}$, where *e* = *2.71828.* This can be done easily with the $\boldsymbol{e^x}$ button on a scientific calculator. Entering -2.3 in such a calculator and pushing the $\boldsymbol{e^x}$ button yields 0.100, or one tenth of a murder per 100,000 population. Actually, the true figure for murders in Missoula County in 1979 and 1991 was zero. Lott used 0.1 instead of zero because the natural logarithm of zero is mathematically undefined, so leaving it at zero would have created missing data. There are a great many -2.3's in his data files on murder, because many of the counties are quite small, much smaller than Missoula with 81,904 people in 1992.

Murder Rates in Five Counties Affected by Right to Carry Laws



The distribution of murder rates in American counties is not at all close to the bell-shaped normal curve. There are great many small counties with few or no murders, and a few quite large ones with a great many. Converting the data to natural logarithms is one way of minimizing the statistical effects of non-normal distributions, but it can introduce other distortions as we see in this case.

Leaving aside the distortions in Missoula County caused by the conversion of the data to natural logarithms, the trends in these counties are quite smooth. There is no apparent effect from the introduction of "shall issue" laws in Missoula County in 1991, in Fulton County (Atlanta, Georgia) in 1990, in Hinds County (Jackson, Mississippi) in 1990, in Fairfax County (Fairfax, Virginia) in 1988 and Kanawha County (Charleston, West Virginia) in 1989.

One might ask, why are we dealing with these medium sized counties instead of major population centers? This was my first clue to the fundamental flaw in Lott's argument. My first inclination was to graph the trends in America's largest cities, because that's where the homicide problem is most severe. I immediately discovered that ***none of these cities had a "shall issue" law****.* The "shall issue" laws were put into effect primarily in states with low population density. This meant that Lott's data did not meet the fundamental assumptions for a regression analysis. To work properly, multiple regression requires that the "shall issue" variable be normally distributed throughout the data set. The mathematical calculations used to "control" for spurious relationships can't work if there is not a sufficient range of variation in the key variables. This was the "smoking gun" hidden in Lott's mass of tables and sophisticated equations. At no point in the book did he acknowledge this fact. When I asked him about this, he shrugged it off. He didn't did not see it as a problem, since he "controlled" for population size.

These irregularities in the data completely invalidated Lott's analysis. It took two years before Ayres and Donohue (1999) verified this in an econometric analysis, but Zimring and Hawkins zeroed in immediately on the problem in 1997. Having studied gun control legislation, 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 (Zimring and Hawkins 1997: 50) "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 (1977: 51) 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.*

What Lott and Mustard were doing was 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, most of them among people who were quite well armed despite the lack of gun permits. 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.

Why Regression Fails: Lack of Predictive Testing.

The acid test in statistical modeling is prediction. A useful model should make predictions that are better than random guessing. Only in this way can cause and effect be distinguished and causal predictions be tested. Regression modelers often do this with historical data, in effect using data from the more distant past to predict the more recent past. The problem with this is that, when the outcome is already known, it is too easy to adjust the model to fit the known outcome. This is like using the day before yesterday's weather to predict yesterday's weather, or the day before yesterday's stock prices to predict yesterday's prices. This may be useful as a way of learning, but the only really persuasive test is to predict tomorrow's weather or stock prices. This criterion, success in prediction, is used to evaluate models of financial markets, the weather, medical outcomes, population trends and many other phenomena. These models all work imperfectly, and regression gives us a good measure of just how imperfectly.

Unfortunately, researchers who use econometric techniques to evaluate social policies generally do not subject their models to predictive tests. They could do so, either by making future predictions and waiting to see what happens, or, if that would take too long, by developing their model with data from one population and then using it to predict results with another population. But most researchers simply do not do this, or if they do the models do not work so the results are never published. (The Urban Institute did make a prediction, but they did not wait for the outcome data to publicize their conclusions. When the data showed that their model did not work, they simply took it down from their WEB site.)

The journals that publish these studies do not require predictive testing, which suggests that the editors and reviewers have low aspirations for their fields. Instead, researchers take data for a fixed period of time and keep fine tuning and adjusting their model it until they can "explain" the 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 social science journals and helps young professors get tenure. Everybody pretends not to notice that little or no progress is being made.

The Alternative: Insist on Intelligible Graphs and Tables

When junk science is released to the media by scholars at prestigious universities, and published in reputable refereed journals, people become understandably skeptical about the value of social science research. A few years ago **The Economist** (May 13, 1995) published a tongue-in-cheek editorial announcing that “74.6% of sociology is bunk.” Cynics wondered if the estimate might not have been low. But it is important not to throw out the baby with the bath water. There is good solid work being done in sociology, criminology and other social sciences, although it may not make it into the journals that value statistical complexity over reliable findings. The most reliable work uses simpler statistical techniques that do not require so much adjustment and standardizing of the data. This has the great advantage that it the work can be read and used by people who have not devoted years of their lives to learning obscure econometric techniques.

Studies that make extensive use of graphics, such as those of Sellin (1959) and Blumstein and Wallman (2000) have been much more successful and informative than regression studies. As an example of the power of simple graphical techniques, we can graph some of the data from John Lott's gun control data set. When a data set is so huge, it may be necessary to select a small part of it for graphing, but this can be quite informative if the selection is done well. In reviewing Lott's data, I found that in the state of Pennsylvania, a "shall issue" law was passed in 1989, but the city of Philadelphia was exempted from it. This provided an excellent opportunity for "natural experiment," comparing trends in two metropolitan areas that differed on a key variable. The graph that follows compares trends in Philadelphia, which is a city and a county, with those in Allegheny County, which includes Pittsburgh. The graph shows that murder rates are generally higher in Philadelphia than in Allegheny County, but the passage of a law giving citizens the right to get permits to carry concealed weapons did not have the positive effect posited by John Lott. In fact, the Allegheny County murder rate was declining prior to the passage of the law, then increased slightly. In Philadelphia, the murder rate had been increasing, then it leveled off despite the fact that the new law did not apply in that city. The violent crime statistics for the same two counties show the same pattern. Disaggregating the data in this way allows us to draw on our qualitative, historical knowledge in interpreting statistical trends. To discredit this kind of finding, concealed weapons advocates would have to show how other factors somehow compensated for the failure of the shall issue law to have any apparent effect.

Conclusions.

These cases may be enough to persuade most readers that multiple regression is not of much use in proving causal arguments, at least about the historical impact of social policies. In fact, the problem is broader than that, and many specialists doubt that multiple regression is a valid way of settling any kind of theoretical argument. In 1985, Stanley Lieberson (1985: ix), a distinguished professor at the University of California, wrote “I am fully sympathetic with the empirical research goals found in much of contemporary sociology, with its emphasis on rigor and quantification. However…I have reluctantly reached the conclusion that many of the procedures and assumptions in this enterprise are of no more merit than a quest for a perpetual motion machine.” 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 in the social sciences 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.”

The social science community does not have good procedures for acknowledging the failure of a widely used research method. Methods that are entrenched in graduate programs at leading universities and published in prestigious journals tend to be perpetuated. Many laymen assume that if a study has been published in a good, peer reviewed journal, there is some assurance of its validity. The cases we have examined here show that this is not the case. Peer review assures that established practices have been followed, but it is of little help when those practices themselves are faulty.

Finding the flaws in regression studies is difficult. Usually, the only way to be sure of them is to obtain the data set and reanalyze the data. This is more than can be expected of a reviewer from a professional journal. It takes time, usually a year or two, for a multiple regression study to be replicated, and many studies never get replicated because they are not of sufficient interest to anyone. Even if a replication is done and does not confirm a study, journal editors may feel that this is simply a case of normal scientific debate. The problem, however, is that no real progress occurs. We are no closer to having a useful mathematical model for predicting homicide rates than we were when Ehrlich published his paper in 1975.

There are no important findings in sociology or criminology that are so complex that they cannot be communicated with graphs and tables that are intelligible to intelligent laymen and policy makers. It is time to admit that the emperor has no clothes. Multiple regression and other mathematical modeling techniques have simply not lived up to their promise as a method of evaluating the impact of public policies. Studies that claim otherwise are junk science.

References

Ayres, I. and Donohue, J. (1999). "Nondiscretionary Concealed Weapons Laws: A Case Study of Statistics, Standards of Proof, and Public Policy." **Am. Law and Ec. Revv** 1: 436-470.

Berk, R.A. (1991) "Toward a Methodology for Mere Mortals," **Sociological Methodology** 21: 315-324.

Black, D. and Nagin, D. (1998). "Do Right-to-Carry Laws Deter Violent Crime?" **J. Legal Studies** 27: 209-219.

Blumstein, A. and Wallman, J. (eds.), **The Crime Drop in America,** Cambridge University Press, New York, pp. 13-44.

Boldt, D. (1999). "Study Evidence on Guns," **Philadelphia Inquirer**, December 14. Downloaded on May17, 2000 from

http://www.phillynews.com/inquirer/99/Dec/14/opinion/BOLDT14.htm.

Bonner, R. and Fessendren, F. (2000). "States with No Death Penalty Share Lower Homicide Rates," **New York Times**, September 22. Downloaded from:

http://www.nytimes.com/2000/09/22/national/22DEAT.html.

Donohue, J. and Levitt, S. (1999). "Legalized Abortion and Crime." Stanford University Law School. Downloaded in August, 2000 from:

http://papers.ssrn.com/paper.taf?ABSTRACT_ID=174508.

Ehrlich, I. (1975). "The Deterrent Effect of Capital Punishment: A Question of Life and Death," **Am. Ec. R.** 65: 397-417.

Fox, J. (2000). "Demographics and U.S. Homicide," In Blumstein, A. and Wallman, J. (eds.), **The Crime Drop in America,** Cambridge University Press, New York, pp. 288-317.

Freedman, D.A. (1991) "Statistical Models and Shoe Leather." **Sociological Methodology** 21: 291-313.

Lott, J and Mustard, D. (1997). "Crime, Deterrence and the Right to Carry Concealed Handguns," **J. Legal St.** 26: 1-68. Downloaded on August 10, 2000 from:

http://www.journals.uchicago.edu/JLS/lott.pdf

Lott, J. (2000). **More Guns, Less Crime: Understanding Crime and Gun Control Laws.** University of Chicago Press, second edition with additional analyses.

Lott, J. and J. Whitley, "Abortion and Crime: Unwanted Children and Out-of-Wedlock Births," Yale Law & Economics Research Paper No. 254. Downloaded on July 9, 2001 from

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=270126.

Manski, C. (1978). "Prospects for Inference on Deterrence through Empirical Analysis of Individual Criminal Behavior." In Blumstein, C, et. al, eds, **Deterrence and Incapacitation,** Washngton: National Academy of Sciences, pp. 400-424.

Marvell, T. and Moody, C. (1997). "The Impact of Prison Growth on Homicide." **Homicide St.** 1: 205-233.

Sellin, T. (1959). **The Death Penalty.** American Law Institute, Philadelphia.

Zimring, F. and Hawkins, G. (1997). "Concealed Handguns: The Counterfeit Deterrent**," The Responsive Community** 7: 46-60.

Econometric modeling and multiple regression
(Letters to the Editor)

Skeptical Inquirer, May, 2002

http://www.findarticles.com/cf_0/m2843/3_26/85932636/p1/article.jhtml

While Ted Goertzel's article "Myths of Murder and Multiple Regression" (January/ February 2002) raises a number of important issues, I fear that its treatment of them does more harm than good. In addition to terminological difficulties ("econometric modeling," "structural equation modeling," "path analysis," and "multiple regression" are not synonymous!), Goertzel's article falsely implies that modern quantitative data analysis (much or all of which he seems to equate with multiple regression) is simply an "arbitrary" game with little connection to reality.

To put it bluntly, it is simply not true that competent data analysts "can achieve any results they want without violating the rules of regression analysis in any way," as many social science articles amply attest. Indeed, the highly restrictive "rules" of OLS render it inappropriate for many problems, often prompting conscientious researchers to employ alternative approaches. The charge that such modeling is instead motivated by some sort of "statistical one-upmanship" is nonsense. Rather, researchers are responding to the need to use more sophisticated techniques (e.g., multivariate time series methods) in order to address questions such as causality (which, Goertzel correctly notes, cannot be directly inferred from multiple regression).

Blind reliance on multiple regression per se is certainly foolish, but poorly aimed criticisms such as those in "Myths of Murder and Multiple Regression" do little to help matters. In my opinion, Goertzel's time might be better spent promoting the use of robustness testing, repeated subset cross-validation, and other well-known strategies for assessing expected predictive power on the basis of limited data.

Carter T. Butts
Department of Social and Decision Sciences
Carnegie Mellon University
Pittsburgh, Pennsylvania

Though I share many of Goertzel's concerns and applaud his call for truly predictive tests, I believe that his basic criticism may have been misdirected and that statistical predictions are in fact underutilized in many important decision-making contexts.

Any statistical procedure can be misused or abused, and results must be interpreted in light of methodology (e.g., research design, sampling, measurement reliability and validity, mathematical assumptions). Indeed, there were sufficient methodological grounds to question the validity of the results in each of the four cited examples. Rather than blaming a statistical tool, the researchers themselves must be held accountable for their oversights (not to mention the editors and reviewers at journals that published their work; those who trust Internet sources that lack rigorous peer review always do so at their own peril).

A substantial research literature demonstrates the superiority of statistical decision making over unaided clinical judgment. Although statistical predictions are far from perfect, their empirical accuracy routinely exceeds that of experts in many professions (e.g., medical/psychological diagnosis, parole hearings, college/graduate admissions, personnel decisions), often by a substantial margin. Despite this impressive track record, there remains a preference for using one's head to integrate information. Significant improvements in accuracy and efficiency may remain unrealized until more people overcome their general distrust of statistical predictions and embrace them in areas where they have demonstrable predictive utility.

John Ruscio
Department of Psychology
Elizabethtown College
Elizabethtown, Pennsylvania

While I appreciate Ted Goertzel's skeptical outlook toward the misuse of statistical techniques and its potentially dangerous outcomes in terms of social policy, I would like to strike a chord of balance concerning statistical analyses in the social (and other) sciences.

Goertzel repeats the statistician's mantra that "correlation is not causation," but this phrase is often interpreted to mean that there is no relationship whatsoever between causation and correlation. If that were true, not only the social sciences, but also much of biology, geology, chemistry, and physics- not to mention statistics itself-should go the way of the phlogiston.

As Bill Shipley has nicely pointed out in his book Cause and Correlation in Biology (Cambridge 2000) a better way to think of the problem is by using a metaphor based on wayang kulit. In this form of oriental art, objects are used to cast a shadow on a screen and the spectator only sees the shadows, interpreting them according to the context of the play. The real objects causing the shadows are always hidden to the observer. The relationship between causes and correlations is pretty much the same: phenomena and forces in the world that are causally connected generate shadows in the form of correlations. Since many different objects can cast what appears to be the same shadow, it is difficult to go from correlations to the underlying causation. This is what philosophers often refer to as the problem of "underdetermination of theories by the data."

Now, as Shipley points out and Goertzel neglects to, some statistical analyses are better than others at establishing a connection between shadows and objects. Multiple regression-criticized by Goerrzel-is indeed one of the worst approaches. But other techniques casually mentioned in the SI article, such as structural equation modeling, can actually be used to unveil some of the correspondences between correlations and their underlying causes, if used properly and with carefully collected data.

While not disagreeing with the spirit of Goertzel's article, I fear that most readers will come away with the impression that statistics is, as Mark Twain put it, "one damn lie after another." That would be a disservice to science and critical thinking.

Massimo Pigliucci
Association Professor of Ecology & Evolutionary Biology
University of Tennessee
Knoxville, Tennessee

Ted Goertzel's article on econometric (and other types of mathematical) modeling was interesting, internally consistent, but also flawed in its basic analytical assumptions and, therefore, in the conclusions reached. His thesis that it is "junk science" to use models "with no demonstrated predictive capability" to deal with complex social issues misses the intent, and thus the value of the work that he so severely criticizes.

As a sociologist, he is surely aware that we do nor now have the capability to identify and quantify all the factors that affect such complex systems.

Indeed, there is not real evidence to support the proposition that these systems can ever be accurately represented by predictive models. Nonetheless, we still must deal with these situations.

These models are an attempt to develop an explicit description (theory) of at least some of the factors critical to the operation of a specific system (question) and how they interact to effect what happens. The result is an analytical tool for continued study and research, not a predictive model. The theory is presented in sufficient detail so that other investigators have the opportunity to further pursue the question. When the model (theory) is modified and run again, and a new result obtained, we learn more about the identified factors and their interactions. The charge that these model exercises are not real science and that there is "little or no progress" made is not justified when viewed from this perspective. Those doing this work are following true scientific principles: they establish a theory which can be tested by others, and which can be used to identify valuable questions for further research.

Unfortunately, model results are often used by partisan advocates to advance their particular ideology. Often they do not really understand the work they are quoting, and misuse the results terribly. This is, I believe, where the junk science comes in. We must distinguish between those who are trying to understand complex social issues and how we might deal with them, and others who seek only to advance a particular viewpoint that they have taken on faith alone. This is an area that has not been covered in SI, and one that, perhaps, should be. It would be interesting to have another article from Mr. Goertzel written from this perspective.

Robert L. Folstein
Germantown, Maryland

I would have been more impressed if Goertzel had included faulty research used by the anti-gun "Brady Bunch" instead of only dogma of the National Rifle Association. An "econometric analysis" (often quoted by Bradyites) of homicides in Seattle and Vancouver claims that gun control caused Vancouver's lower homicide rate because the two are demographically similar. Even though the latter has few blacks or Hispanics, Seattle whites have a lower homicide rare than Vancouver whites, and the difference in homicide rates is strongly correlated to Seattle's black and Hispanic population!

Is multiple regression difficult to use correctly? And how! Lott controlled for hundreds of variables and concluded a change of a few percentage points. How about sensitivity analysis for the controlled variables? Is the result believable without knowing that a minor error in a controlled variable could make the result vastly different? Or is that Lott's incomprehensible complexity? Is Goertzel possibly nor competent to criticize?

Good statistical analysis is very difficult. That should have been Goertzel's explicit and implicit point.

If SKEPTICAL INQUIRER is part of the group of media believing Bradyitcs correct and NRA wrong without rigorous study of conflicting data, I will cancel my subscription. Skeptical means skeptical, no matter who!

Sam Brunstein
Burbank, California

The question of gun control is intensely political, and statements on one side should be viewed just as skeptically as those on the other side.

Goertzel cites objections by Zimring and Hawkins to the work of Lott. It would be appropriate to note that Zimring and Hawkins have long preached essentially prohibitionist firearms control, and can hardly be considered as neutral observers. They seem to me to be religiously defending doctrine against observation. Lott, on the other hand, had apparently not previously been involved in the gun control controversy at all. That is consistent with his own statement that he became interested only because of questions from a student in one of his classes. Goertzel quotes Zimring and Hawkins as saying that "a determined econometrician can produce a treatment of the same historical periods with different models and opposite effects." He then refers to a report by Black and Nagin, whom he considers those determined econometricians: ". . . if they changed the statistical model a little bit, or applied it to different segments of the data, Lott and Mustard's findings disappeared." However, the little bit of change amounted to deletion of most of the data. (Florida and all counties with less than 100,000 population, or about 85% of all United States counties). That left a relatively small sample loaded with heavily populated areas with high crime and very few shall-issue laws. Even then, they did not report a harmful effect of "shall issue" laws, but only that the reduction in crime was not statistically significant for the relatively few counties, with relatively little variation in shall-issue laws, on which they report.

Goerrzel says that Lott's analysis is unsuitable because, in covering all the counties in the United States, it includes large cities which do not have "shall issue" laws and which do have particularly high incidences of murder. In the next paragraph he directly contradicts that, saying that Lott "simply had no data for the major cities where the homicide problem was not acute." Politically incorrect science is not at all the same thing as junk science.

Denzel L. Dyer
Rancho Palos Verdes, California

Ted Goertzel's article unwittingly serves as a good example of how ideology may bias one's judgment.

Take Goertzel's "Myth One": If John Lott's finding, that "right to carry" gun laws lower rates of violent crime, were merely an artifact of how he chose to analyze the data, then opponents of gun ownership should be able to use the same data to show such laws cause higher crime rates. According to Goertzel, however, Lott's opponents had to remove the entire state of Florida from the data; and this did nor reverse Lott's findings, but merely resulted in "no detectable impact on the rate of murder and rape." As guncontrol advocates had earlier predicted bodies piling up in the streets, clearly Lott's findings have proven robust enough to turn the debate around, if nor to end it.

"Myth Three": Goertzel attacks Isaac Ehrlich's famous research on American executions and murder deterrence. He does not mention the recent research on British executions that produced almost exactly the same result: an execution deters eight or more murders.

Goertzel urges us to demand that models demonstrate their predictive ability; but he never mentions the inverse correlation between U.S. executions and murders in the second half of the twentieth century: as executions were reduced to zero, the murder rate soared; as executions resumed, the murder rate fell. Capital punishment opponents had once predicted the opposite: banning executions was supposed to teach criminals respect for human life, by example! But once again the debate has been turned around, and capital punishment opponents are left insisting, merely, that there is "no demonstrable effect."

Taras Wolansky
Kerhonkson, New York

Professor Goertzel's expose of some arcane statistical methods appears to be left-wing politics crossdressing as analysis of econometric modeling and multiple regression.

For example, in Lott and Mustard's book More Guns. Less Crime, straightforward data indicate, even without any fancy statistical footwork, that laws passed allowing honest citizens to arm themselves against criminals reduces crime. No great surprise there, except that this common-sense finding demolishes the implied presumptions of fervid gun control advocates.

In the U.S., most states, admittedly not the most populous nor liberal ones, have passed these shall-carry laws in order to permit ordinary citizens, without criminal records or mental problems, to carry concealed guns. These laws, even as Goertzel allows, have been shown to be beneficial and to have saved lives and prevented robberies, rapes, and other acts of violence by the thousands, where they have been enacted.

One would think even Goertzel and guncontrol advocates would find this virtuous correlation intriguing and wonder if the same principle would work in populous states, instead of making complaints about exotic statistics. The fact that they do not might make one suspect them of ideological junk analysis.

Don Vandervelde
Gig Harbor, Washington

I have enjoyed the latest issue of your fine magazine, but noted an error in the article "Myths of Murder and Multiple Regression" on page 20. The statement: "Lott had collected data for each of America's 50,056 counties for each year from 1977 to 1992" is grossly inaccurate. There were only 3,076 counties in America at that time, counting the Judicial Districts in Alaska, Parishes in Louisiana and counties of the Islands of Hawaii. A new county has been added by Colorado recently to make the current total 3,077.

Harry H. Incho
Medina, New York

Ted Goertzel replies:

John Ruscio is correct that statistical predictions are often superior to clinical judgments, for reasons that he explained quite well in his own article on "The Emptiness of Holism" in the March/April SKEPTICAL INQUIRER. The key word here is prediction. I'm entirely in favor of statistical models that predict trends in data, other than the data used to build the models. The techniques suggested by Carter Butts may be useful for developing good predictive models. If he can get them to work, more power to him. My quarrel is with using multiple regression to "prove" causal arguments. This has done a lot of harm, beyond the examples I cited in my paper. A good example is the very controversial book The Bell Curve by Herrnstein and Murray. It used multiple regression to make arguments about the causes of the black-white test score gap. Clark Glymour, a distinguished philosopher of science, observed that "the unstated problem for many commentators is how to reject the particular conclusions of The Bell Curve without also rejecting the larger enterprises of statistical social science...." It can't be done, except by rejecting multiple regression as a means of proving causal arguments. Herrnstein and Murray's multiple regression modeling techniques are just as good as those used by other researchers who reach opposite conclusions. Anyone wanting a more technical, philosophical explanation of why multiple regression cannot answer this kind of question should read Glymour's chapter "Social Statistics and Genuine Inquiry" in Intelligence, Genes, and Success: Scientists Respond to **The Bell Curve** published by Springer in 1997.

Another example of a faulty model was the Urban Institute's prediction that a million children "might" be thrown into poverty by the 1996 welfare reform. This kind of junk science undermines public confidence in social science altogether. The important thing is not the particular technique, it is whether the models have demonstrated predictive validity. There are, as Carter Butts states, technical differences between modeling techniques. But these are not relevant to the points I was making in my article. My advice to statistical modelers is simple: no more training and testing on the same data!

I agree with Massimo Pigliucci and Robert Folstein that statistical modeling has its uses, and that much of the problem is the misuse of results by advocates. But it is not only journalists and activists who misuse the results. The Urban Institute actively promoted its predictions in an attempt to stop welfare reform. Publishing a book with the title More Guns, Less Crime based on the data John Lott had was simply irresponsible. Don Vandervelde's hunch is understandable based on the analyses Lott published, but he is factually wrong: crime did not go down in the counties that passed shall-carry laws. But you would never know this from reading Lott's book. I know it because I got a copy of Lott's data set and did a number of county-by-county trend analyses myself. There was no room for these graphs in the short paper I published in the SKEPTICAL INQUIRER, but they are available in a longer paper available for download on my Web site: http://goertzel.org/ted. There is also a discussion of the Urban Institute study on the Web site.

There are many other articles I might have reviewed. My focus was on the misuse of multiple regression modeling, not on gun control or the death penalty per se. The problems with Seattle/Vancouver study Sam Brunstein mentions were not with multiple regression but a failure to consider the importance of race in explaining the different trends in the cities. Rather than focusing on one causal variable, it is better to focus on a dependent variable, such as homicide rates, and then examine all the factors that may explain the trend. There are many factors that explain the trends in homicide rates in the United States in the second half of the twentieth century better than the correlations with execution rates that Taras Wolansky points to.

Harry Incho is right about the number of counties in the United States. Lott had 50,076 county level measurements, counting each year's data as a separate measurement. Denzel Dyer misunderstood my argument; my point was that Lott had no data on his causal variable--shall-issue laws--for the major cities. He nevertheless threw them into the multiple regression calculations, which distorted his results. If he had left the large cities out it would have been better.

Finally, I like Robert Folstein's suggestion that I write an article on how to distinguish honest attempts at untangling difficult issues from one-sided advocacy research. I do not share his optimism about the usefulness of complex modeling in fields where there is not sufficient data for predictive testing. But I would welcome examples of studies where it has been helpful, as well as of cases where it has been misused. Send your examples to me at

goertzel@camden.rutgers.edu.