AMERICAN
Scientist

## Freakonomics: What Went Wrong?

Examination of a very popular popular-statistics series reveals avoidable errors

Andrew Gelman, Kaiser Fung

The nonfiction publishing phenomenon known as *Freakonomics* has passed its sixth anniversary. The original book, which used ideas from statistics and economics to explore real-world problems, was an instant bestseller. By 2011, it had sold more than four million copies worldwide, and it has sprouted a franchise, which includes a bestselling sequel, *SuperFreakonomics;* an occasional column in the *New York Times Magazine;* a popular blog; and a documentary film. The word "freakonomics" has come to stand for a light-hearted and contrarian, yet rigorous and quantitative, way of looking at the world.
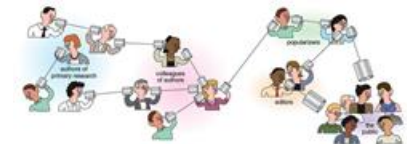
+ enlarge image

The faces of Freakonomics are Steven D. Levitt, an award-winning professor of economics at the University of Chicago, and Stephen J. Dubner, a widely published New York–based journalist. Levitt is celebrated for using data and statistics to solve an array of problems not typically associated with economics. Dubner has perfected the formula for conveying the excitement of Levitt's research—and of the growing body of work by his collaborators and followers. On the heels of *Freakonomics*, the pop-economics or pop-statistics genre has attracted a surge of interest, with more authors adopting an anecdotal, narrative style.

As the authors of statistics-themed books for general audiences, we can attest that Levitt and Dubner's success is not easily attained. And as teachers of statistics, we recognize the challenge of creating interest in the subject without resorting to clichéd examples such as baseball averages, movie grosses and political polls. The other side of this challenge, though, is presenting ideas in interesting ways without oversimplifying them or misleading readers. We and others have noted a discouraging tendency in the Freakonomics body of work to present speculative or even erroneous claims with an air of certainty. Considering such problems yields useful lessons for those who wish to popularize statistical ideas.

On a Case-by-case Basis

In our analysis of the Freakonomics approach, we encountered a range of avoidable mistakes, from back-of-the-envelope analyses gone wrong to unexamined assumptions to an uncritical reliance on the work of Levitt's friends and colleagues. This turns accessibility on its head: Readers must work to discern which conclusions are fully quantitative, which are somewhat data driven and which are purely speculative.

The case of the missing girls: Monica Das Gupta is a World Bank researcher who, along with others in her field, has attributed the abnormally high ratio of boy-to-girl births in Asian countries to a preference for sons, which manifests in selective abortion and, possibly, infanticide. As a graduate student in economics, Emily Oster (now a professor at the University of Chicago) attacked this conventional wisdom. In an essay in *Slate,* Dubner and Levitt praised Oster and her study, which was published in the *Journal of Political Economy* during Levitt's tenure as editor:

> [Oster] measured the incidence of hepatitis B in the populations of China, India, Pakistan, Egypt, Bangladesh, and other countries where mothers gave birth to an unnaturally high number of boys. Sure enough, the regions with the most hepatitis B were the regions with the most "missing" women. Except the women weren't really missing at all, for they had never been born.

Oster's work stirred debate for a few years in the epidemiological literature, but eventually she admitted that the subject-matter experts had been right all along. One of Das Gupta's many convincing counterpoints was a graph showing that in Taiwan, the ratio of boys to girls was near the natural rate for first and second babies (106:100) but not for third babies (112:100); this pattern held up with or without hepatitis B.

In a follow-up blog post, Levitt applauded Oster for bravery in admitting her mistake, but he never credited Das Gupta for her superior work. Our point is not that Das Gupta had to be right and Oster wrong, but that Levitt and Dubner, in their celebration of economics and economists, suspended their critical thinking.

The risks of driving a car: In *SuperFreakonomics*, Levitt and Dubner use a back-of-the-envelope calculation to make the contrarian claim that driving drunk is safer than walking drunk, an oversimplified argument that was picked apart by bloggers. The problem with this argument, and others like it, lies in the assumption that the driver and the walker are the same type of person, making the same kinds of choices, except for their choice of transportation. Such all-else-equal thinking is a common statistical fallacy. In fact, driver and walker are likely to differ in many ways other than their mode of travel. What seem like natural calculations are stymied by the impracticality, in real life, of changing one variable while leaving all other variables constant.

Stars are made, not born—except when they are born: In 2006, Levitt and Dubner wrote a column for the *New York Times Magazine* titled "A Star Is Made," relying on the research of Florida State University psychologist K. Anders Ericsson, who believes that experts arise from practice rather than innate talent. It begins with the startling observation that elite soccer players in Europe are much more likely to be born in the first three months of the year. The theory: Since youth soccer leagues are organized into age groups with a cutoff birth date of December 31, coaches naturally favor the older kids within each age group, who have had more playing time. So far, so good. But this leads to an eye-catching piece of wisdom: The fact that so many World Cup players have early birthdays, the authors write,

> may be bad news if you are a rabid soccer mom or dad whose child was born in the wrong month. But keep practicing: a child conceived on this Sunday in early May would probably be born by next February, giving you a considerably better chance of watching the 2030 World Cup from the family section.

Perhaps readers are not meant to take these statements seriously. But when we do, we find that they violate some basic statistical concepts. Despite its implied statistical significance, the size of the birthday effect is very small. The authors acknowledge as much three years later when they revisit the subject in *SuperFreakonomics.* They consider the chances that a boy in the United States will make baseball's major leagues, noting that July 31 is the cutoff birth

date for most U.S. youth leagues and that a boy born in the United States in August has better chances than one born in July. But, they go on to mention, being born male is "infinitely more important than timing an August delivery date." What's more, having a major-league player as a father makes a boy "*eight hundred times* more likely to play in the majors than a random boy," they write. If these factors are such crucial determinants of future stardom, what does this say about their theory that a star is made, not born? Practice may indeed be a more important factor than innate talent, but in opting for cute flourishes like these, the authors venture so far from the original studies that they lose the plot.

Making the majors and hitting a curveball: In the same discussion in *SuperFreakonomics,* Levitt and Dubner write:

> A U.S.-born boy is roughly 50 percent more likely to make the majors if he is born in August instead of July. Unless you are a big, big believer in astrology, it is hard to argue that someone is 50 percent better at hitting a big-league curveball simply because he is a Leo rather than a Cancer.

But you don't need to believe in astrology to realize that the two cited probabilities are not the same. A .300 batting average is 50 percent better than a .200 average. In such a competitive field, the difference in batting averages between a kid who makes the majors and one who narrowly misses out is likely to be a matter of hundredths or even thousandths of a percent. Such errors could easily be avoided.

Predicting terrorists: In *SuperFreakonomics*, Levitt and Dubner introduce a British man, pseudonym Ian Horsley, who created an algorithm that used people's banking activities to sniff out suspected terrorists. They rely on a napkin-simple computation to show the algorithm's "great predictive power":

> Starting with a database of millions of bank customers, Horsley was able to generate a list of about 30 highly suspicious individuals. According to his rather conservative estimate, at least 5 of those 30 are almost certainly involved in terrorist activities. Five out of 30 isn't perfect—the algorithm misses many terrorists and still falsely identified some innocents—but it sure beats 495 out of 500,495.

The straw man they employ—a hypothetical algorithm boasting 99-percent accuracy—would indeed, if it exists, wrongfully accuse half a million people out of the 50 million adults in the United Kingdom. So the conventional wisdom that 99-percent accuracy is sufficient for terrorist prediction is folly, as has been pointed out by others such as security expert Bruce Schneier.

But in the course of this absorbing narrative, readers may well miss the spot where Horsley's algorithm also strikes out. The casual computation keeps under wraps the rate at which it fails at catching terrorists: With 500 terrorists at large (the authors' supposition), the "great" algorithm finds only five of them. Levitt and Dubner acknowledge that "five out of 30 isn't perfect," but had they noticed the magnitude of false negatives generated by Horsley's secret recipe, and the grave consequences of such errors, they might have stopped short of hailing his story. The maligned straw-man algorithm, by contrast, would have correctly identified 495 of 500 terrorists.

This unavoidable tradeoff between false positive and false negative errors is a well-known property of all statistical-prediction applications. Circling back to check all the factors involved in the problem might have helped the authors avoid this mistake.

The climate-change dustup: Rendering research conducted by others is much more challenging than explaining your own work, especially if the topic lies outside your domain of expertise. The climate-change chapter in *SuperFreakonomics* is a case in point. In it, Levitt and Dubner throw their weight behind geoengineering, a climate-remediation concept championed at the time by Nathan Myhrvold, a billionaire and former chief technology officer of Microsoft. Unfortunately, having moved outside the comfort zone of his own research, Levitt is in no better a position to evaluate Myhrvold's proposal than we are.

When an actual expert, University of Chicago climate scientist Raymond Pierrehumbert, questioned the claims in Levitt and Dubner's writing on climate, Levitt retorted that he enjoyed Pierrehumbert's "intentional misreading" of the chapter. Referring to his own writings on the subject, Levitt wrote, "I'm not sure why that is blasphemy." We're not sure on this point either—we could not find a place where Pierrehumbert described Levitt's writings in those terms. It is easy to be preemptively defensive of one's own work, or of researchers whose work one has covered. Viewing alternative points of view as useful rather than threatening can help take the sting out of critiques. And if you're covering subject matter outside your expertise, it pays to get second—and third and fourth —opinions.

Problems—and Solutions

How could an experienced journalist and a widely respected researcher slip up in so many ways? Some possible answers to this question offer insights for the would-be pop-statistics writer.

Leave friendship at the door: We attribute many of these errors to the structure of the authors' collaboration, which, from what we can tell, relies on an informal social network that has many potential failure points. In the original *Freakonomics*, much of whose content appeared originally in columns for the *New York Times Magazine,* the network seems to have been more straightforward: Levitt did the research, Dubner trusted Levitt, the *Times* trusted Dubner, and we the readers trusted the *Times*'s endorsement. In *SuperFreakonomics* and the authors' blog, it becomes less clear: Levitt trusts brilliant stars such as Myhrvold or Oster, Dubner trusts Levitt, and we the readers trust the Freakonomics brand. A more ideal process for science writing (as shown in the illustration above) will likely look much messier—but it offers the promise of better results.
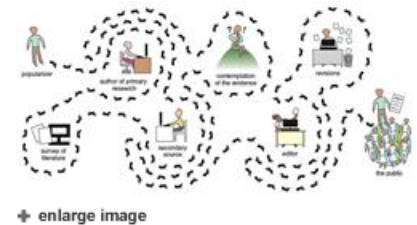
+ enlarge image

Don't sell yourself short: Perhaps Levitt's admirable modesty—he has repeatedly attributed his success to luck and hard work rather than genius—has led him astray. If he feels he is surrounded by economists more exceptional and brilliant than he is, he may let their assertions stand without challenge. Here it might be good to remember the outsider's perspective so prized by Levitt: If you find yourself hesitant to ask questions that seem "stupid," or if you feel intimidated, think of yourself as a "rogue." Just don't take it so far that you value your own rogueness over empirical evidence.

Maintain checks and balances: A solid collaboration requires each side to check and balance the other side. Although there's no way we can be sure, perhaps, in some of the cases described above, there was a breakdown in the division of labor when it came to investigating technical points. The most controversial statements are the most likely to be mistaken; if such assertions go unchallenged, you will have little more than a series of press releases linked by gung-ho commentary and eye-popping headlines. Hiring a meticulous editor who can evaluate the technical arguments is another way to avoid embarrassing mistakes.

Take your time: Success comes at a cost: The constraints of producing continuous content for a blog or website and meeting publisher's deadlines may have adverse effects on accuracy. The strongest parts of the original *Freakonomics* book revolved around Levitt's own peer-reviewed research. In contrast, the Freakonomics blog features the work of Levitt's friends, and *SuperFreakonomics* relies heavily on anecdotes, gee-whiz technology reporting and work by Levitt's friends and colleagues. Just like good science, good writing takes time. Remembering this can help hedge against the temptation to streamline arguments or narrow the pool of sources, even in the face of deadlines.

Be clear about where you're coming from: Levitt's publishers, along with Dubner, characterize him as a "rogue economist." We find this odd: He received his

Ph.D. from the Massachusetts Institue of Technology, holds the title of Alvin H. Baum Professor of Economics at the University of Chicago and has served as editor of the mainstream *Journal of Political Economy.* He is a research fellow with the American Bar Foundation and a member of the Harvard Society of Fellows, and has worked as a consultant for Corporate Decisions, Inc. One can be an outsider within such institutions, of course. But much of his economics is mainstream. And his statistical methods are conventional (which, we hasten to add, is not a bad thing at all!). One of the pleasures of reading *Freakonomics* is Levitt's knack for finding interesting quantitative questions in obscure corners, such as the traveling bagel salesman and cheating sumo wrestlers. Often such problems have not been extensively studied or even been noticed by others, and in these cases one is hard-pressed to identify any consensus or conventional wisdom. Often, in the authors' writing, the "conventional" and the "rogue" live side by side. Chapter one of *SuperFreakonomics*, for instance, can be viewed either as a clear-eyed quantitative examination of the economics of prostitution, or as an unquestioning acceptance of conventional wisdom about gender roles. In exploring new territory, it's especially important to be plainspoken about where your assumptions come from and what your primary ideas are.

Use latitude responsibly: When a statistician criticizes a claim on technical grounds, he or she is declaring not that the original finding is wrong but that it has not been convincingly proven. Researchers—even economists endorsed by Steven Levitt—can make mistakes. It may be okay to overlook the occasional mistake in the pursuit of the larger goal of understanding the world. But once one accepts this lower standard—science as plausible stories or data-supported reasoning, rather than the more carefully tested demonstrations that are characteristic of Levitt's peer-reviewed research articles—one really has to take extra care, consider all sides of an issue, and look out for false positive results.

The landscape of pop-statistics books grows more varied by the year, and Levitt and Dubner's bestsellers have introduced several new ingredients to the genre. One of the delights of the books and the blog is the authors' willingness to play with ideas and consider alternative explanations. But unquestioning trust in friends and colleagues combined with the desire to be counterintuitive appear in several cases to have undermined their work. They—and anyone who wishes to convey economics and statistics to a popular audience—just need to take the next step and avoid, in any given example, privileging one story over all other possibilities. This may require Levitt to be more skeptical of the research of his friends and colleagues, and Dubner to be more skeptical of Levitt. "Easy read" should not mean "easy write."

And it doesn't even always mean "easy read": Readers should apply the same skepticism to the claims of *Freakonomics* as they would to the much-derided conventional wisdom. We encourage them to revisit these modern-day classics with a skeptical and inquiring mind. And we hope that future works in the pop-statistics genre will continue to impart a sense of the fun and importance of statistical reasoning, while more clearly recognizing the uncertainty and complexity inherent in scientific study of the world.

Bibliography

- Das Gupta, M. 2005. Explaining Asia's "missing women": A new look at the data. *Population and Development Review* 31:529–535.
- DiNardo, J. E. 2006. Freakonomics: Scholarship in the service of storytelling. *American Law and Economics Review* 8:615–626.
- Dubner, S. J., and S. Levitt. 2006. A star is made: The birth-month soccer anomaly. *The New York Times Magazine,* May 7. http://www.nytimes.com/2006/05/07/magazine/07wwln_freak.html.
- Dubner, S. J., and S. Levitt. 2005. The search for 100 million missing women. *Slate,* May 24. http://www.slate.com/articles/business/the_dismal_science/2005/05/the_search_for_100_million_missing_women.html.
- Kahneman, D., and A. Tversky. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–1131.
- Levitt, S. 2008. An academic does the right thing. Freakonomics (blog), May 22. http://www.freakonomics.com/2008/05/22/an-academic-does-the-right-thing/.
- Levitt, S. 2009. Comment on Raymond T. Pierrehumbert, An open letter to Steve Levitt, RealClimate (blog), October 30. http://www.realclimate.org/index.php/archives/2009/10/an-open-letter-to-steve-levitt (comment 47).
- Levitt, S., and S. J. Dubner. 2005. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything.* New York: William Morrow.
- Levitt, S., and S. J. Dubner. 2009. *Superfreakonomics: Global Cooling, Patriotic Prostitutes, and Why Suicide Bombers Should Buy Life Insurance.* New York: William Morrow.
- Oster, E. 2005. Hepatitis B and the case of the missing women. *Journal of Political Economy* 113:1163–1216.
- Pierrehumbert, R. 2009. An open letter to Steve Levitt. RealClimate (blog), October 29. http://www.realclimate.org/index.php/archives/2009/10/an-open-letter-to-steve-levitt.
- Rubinstein, A. 2006. Freak-freakonomics. *Economist's Voice* 3(9), article 7.
- Schneier, B. 2007. How to not catch terrorists. *Forbes*, March 26.

You can find this online at http://www.americanscientist.org/issues/num2/freakonomics-what-went-wrong/1

POWERED BY
eResources