Why ask why?
Forward causal inference and reverse causal questions*

Andrew Gelman† Guido Imbens‡

5 Oct 2013

Abstract

The statistical and econometrics literature on causality is more focused on "effects of causes" than on "causes of effects." That is, in the standard approach it is natural to study the effect of a treatment, but it is not in general possible to define the causes of any particular outcome. This has led some researchers to dismiss the search for causes as "cocktail party chatter" that is outside the realm of science. We argue here that the search for causes can be understood within traditional statistical frameworks as a part of model checking and hypothesis generation. We argue that it can make sense to ask *questions* about the causes of effects, but the answers to these questions will be in terms of effects of causes.

Introductory example: cancer clusters

A map shows an unexpected number of cancers in some small location, and this raises the questions: What is going on? Why are there so many cancers in this place? These questions might be resolved via the identification of some carcinogenic agent (for example, a local environmental exposure) or, less interestingly, some latent variable (some systematic difference between people living in and not living in this area), or perhaps some statistical explanation (for example, a multiple-comparisons argument demonstrating that an apparent anomalous count in a local region is no anomaly at all, after accounting for all the possible clusters that could have been found). The question, Why so many cancers in that location?, until it is explained in some way, refers to a data pattern that does not fit one's pre-existing models. Researchers are motivated to look for an explanation because this might lead to a deeper understanding and ultimately, through the identification of some carcinogenic agent, to recommendations for policy that may lead to a reduction in cancer rates.

The cancer-cluster problem is typical of a lot of scientific reasoning. Some anomaly is observed and it needs to be explained. The resolution of the anomaly may be an entirely new paradigm (Kuhn, 1970) or a reformulation of the existing state of knowledge (Lakatos, 1978). In the cancer example, potential causes might be addressed via a hierarchical regression model as in Manton et al. (1989).

The purpose of the present paper is to place "Why" questions within a statistical framework of causal inference. We argue that a question such as "Why are there so many cancers in this place?" can be viewed not directly as a question of causal inference, but rather indirectly as an identification of a problem with an existing statistical model, motivating the development of more sophisticated statistical models that can directly address causation in terms of counterfactuals and potential outcomes.

*We thank Judea Pearl and several other blog commenters for helpful input and the National Science Foundation for partial support of this work.

†Department of Statistics, Columbia University, New York, N.Y.

‡School of Business, Stanford University, Stanford, Calif.

Forward and reverse causal questions

Following Gelman (2011), we distinguish two broad classes of causal queries:

1. *Forward causal questions*, or the estimation of "effects of causes." What might happen if we do X? What is the effect of some manipulation, *e.g.,* the effect of smoking on health, the effect of schooling on earnings, the effect of campaigns on election outcomes, and so forth?

2. *Reverse causal inference*, or the search for "causes of effects." What causes Y? Why do more attractive people earn more money? Why does *per capita* income vary some much by country? do many poor people vote for Republicans and rich people vote for Democrats? Why did the economy collapse?

When statistical and econometric methodologists write about causal inference, they generally focus on forward causal questions. We are taught to answer questions of the type "What if?", rather than "Why?" Following the work by Rubin (1977) causal questions are typically framed in terms of manipulations: if x were changed by one unit, how much would y be expected to change? But reverse causal questions are important too. They are a natural way to think (consider the importance of the word Why, or see Sloman, 2005, for further discussion of causal ideas in cognition). In many ways, it is the reverse causal questions that motivate the research, including experiments and observational studies, that we use to answer the forward questions.

The question discussed in the current paper is: How can we incorporate reverse causal questions into a statistical framework that is centered around forward causal inference?[1] Our resolution is as follows: Forward causal inference is about estimation; reverse causal questions are about model checking and hypothesis generation. To put it another way, we ask reverse causal *questions* all the time, but we do not perform reverse causal *inference*. We do not try to answer "Why" questions; rather, "Why" questions motivate "What if" questions that can be studied using standard statistical tools such as experiments, observational studies, and structural equation models.

Before going on, let us clarify that we are using the terms "forward" and "reverse" to refer not to time but to the sequence of the statistical model. As Fearon (1991) and Yamamoto (2012) discuss, counterfactual reasoning can also be applied to assess attributions of causes of past events. We label this historical reasoning as "forward causal inference" as well, as it is based on the estimation of effects of defined treatments. For a recent example, Ashraf and Galor (2013) estimate a nonlinear effect of genetic diversity on historical economic performance of countries around the world. This study may well have been *motivated* by a reverse causal question (Why are African and South American countries relatively poor?) and a feeling that such a question was not adequately answered by existing models such as the geographical explanations popularized by Diamond (1997), but once the question has been framed in terms of the effect of a specific variable, its resolution is conducted within the realm of forward causal reasoning. In that sense, macro studies of the causes of variation in

[1] Even methods such as path analysis or structural modeling, which in some settings can be used to determine the direction of causality from data, are still ultimately answering forward casual questions of the sort, What happens to y when we change x?

economic growth are related to micro studies of the effects of particular interventions (Banerjee and Duflo, 2011). The anomalies identified by Why questions motivate experiments and observational studies for forward causal inference and, ultimately, policy recommendations.

Example: political campaigns

An important forward causal question is, What is the effect of money on elections? This is a very general question and, to be answered, needs to be made more precise, for example as follows: Supposing a challenger in a given congressional election race is given a $100 donation, how much will this change his or her expected vote share? It is not so easy to get an accurate answer to this question, but the causal quantity is well defined (and can be specified even more precisely, for example with further details about the contribution, as necessary).

Now a reverse causal question: Why do incumbents running for reelection to Congress get so much more funding than challengers? Many possible answers have been suggested, including the idea that people like to support a winner, that incumbents have higher name recognition, that certain people give money in exchange for political favors, and that incumbents are generally of higher political "quality" than challengers and get more support of all types. Various studies could be performed to evaluate these different hypotheses, all of which could be true to different extent (and in some interacting ways).

Now the notation. It is generally accepted (and we agree) that forward causal inferences can be handled in a potential-outcome or graphical-modeling framework involving a treatment variable T, an outcome y, and pre-treatment variables, x, so that the causal effect is defined (in the simple case of binary treatment) as $y(T = 1, x) - y(T = 0, x)$. The actual estimation will likely involve some modeling (for example, some curve of the effect of money on votes that is linear at the low end, so that a $20 contribution has twice the expected effect as $10), but there is little problem in defining the treatment effect. The challenge arises from estimating these effects from observational data; see, for example, Levitt (1994) and Gerber (1998).

Reverse causal reasoning is different; it involves asking questions and searching for new variables that might not yet even be in our model. We can frame reverse causal questions as model checking. It goes like this: what we see is some pattern in the world that needs an explanation. What does it mean to "need an explanation"? It means that existing explanations—the existing model of the phenomenon—does not do the job. This model might be implicit. For example, if we ask, Why do incumbents get more contributions than challengers, we are comparing to an implicit model that all candidates get the same. If we gather some numbers on dollar funding, compare incumbents to challengers, and find the difference is large and statistically significant, then we are comparing to the implicit model that there is variation but not related to incumbency status. If we get some measure for candidate quality (for example, previous elective office and other qualifications) and still see a large and statistically significant difference between the funds given to incumbents and challengers, then it seems we need more explanation. And so forth.

Further examples

Kaiser Fung, a statistician who works in the corporate world, writes:

> A lot of real world problems are of the reverse causality type, and it's an embarrassment for us to ignore them. . . . Most business problems are reverse causal. Take for example P&G who spend a huge amount of money on marketing and advertising activities. The money is spread out over many vehicles, such as television, radio, newspaper, supermarket coupons, events, emails, display ads, search engine, etc. The key performance metric is sales amount.
>
> If sales amount suddenly drops, then the executives will want to know what caused the drop. This is the classic reverse causality question. Of course, lots of possible hypotheses could be generated . . . TV ad was not liked, coupons weren't distributed on time, emails suffered a deliverability problem, etc. By a process of elimination, one can drill down to a small set of plausible causes. This is all complex work that gives approximate answers.
>
> The same question can be posed as a forward causal problem. We now start with a list of treatments. We will divide the country into regions, and vary the so-called marketing mix, that is, the distribution of spend across the many vehicles. This generates variations in the spend patterns by vehicle, which allows us to estimate the effect of each of the constituent vehicles on the overall sales performance.

This is the pattern: a quantitative analyst notices an anomaly, a pattern that cannot be explained by current models. The reverse-causal question is: Why did this happen? And this leads to improved modeling. Often these models are implicit. For example, consider the much-asked question, Why did violent crime in American cities decline so much in recent decades? This is a question asked in reference to summaries of crime statistics, not with reference to any particular model. But the Why question points to factors that should be considered. Our country has had similar discussions along the lines of, Why was there a financial crisis in 2008, and why did it happen then rather than sooner or later?

Again, the goal of this paper is to discuss how such Why questions fit into statistics and econometrics, not as inferential questions to be answered with estimates or confidence intervals, but as the identification of statistical anomalies that motivate improved models.

Relation to formal models of causal inference

Now let us put some mathematical structure on the problem using the potential outcome framework (Neyman, 1923, Rubin, 1974).

Let Y_i denote the outcome of interest for unit i, say an indicator whether individual i developed cancer. We may also observe characteristics of units, individuals in this case, that are known to, or expected to, be related to the outcome, in this case, related to the probability of developing cancer. Let us denote those by W_i. Finally, there is a characteristic of the units, denoted by Z_i, that the researcher feels should not be a cause of the outcome, and

so one would expect that in populations homogenous in W_i, there should be no association between the outcome and Z_i:

$$Y_i \perp Z_i \mid W_i.$$

This atribute Z_i may be the location of individuals. It may be a binary characteristic, say female versus male, or an indicator for a subpopulation. The key is that the researcher *a priori* interprets this variable as an attribute that should not be correlated with the outcome in homogenous subpopulations.

However, suppose the data reject this hypothesis, and show a substantial association between Z_i and the outcome, even after adjusting for differences in the other attributes. In a graphical representation of the model there would be evidence for an arrow between Z_i and Y_i, in addition to the arrow from W_i and Y_i, and possibly a general association between Z_i and W_i.

Such finding is consistent with two alternative models. First, it may be that there is a cause, its value for unit i denoted by X_i, such that the potential outcome given the cause is not associated with Z_i, possibly after conditioning on W_i. Let $Y_i(x)$ denote the potential outcomes in the potential outcome framework, and $Y_i(X_i)$ the realized outcome. Formally we would hypothesize that, for all x,

$$Y_i(x) \perp Z_i \mid W_i.$$

Thus, the observed association between Y_i and Z_i is the result of a causal effect of X_i on Y_i, and an association between the cause X_i and the attribute Z_i. In the cancer example there may be a carcinogenic agent that is more common in the area with high cancer rates.

The second possible explanation that is still consistent with no causal link between Z_i and Y_i is that the researcher omitted an important attribute, say V_i. Given this attribute and the original attributes W_i, the association between Z_i and Y_i would disappear:

$$Y_i \perp Z_i \mid W_i, V_i.$$

For example, it may be that individuals in this area have a different genetic background that makes them more susceptible to the particular cancer.

Both these alternative models could in principle provide a satisfactory explanation for the anomaly of the strong association between Z_i and the outcome. Whether these models do so in practice, and which of these models do, and in fact whether the original association is even viewed as an anomaly, depends on the context and the researcher. Consider the finding that taller individuals command higher wages in the labor market. Standard economic models suggest that wages should be related to productivity, rather than height, and such an association might therefore be viewed as an anomaly. It may be that childhood nutrition affects both adult height and components of productivity. One could view that as an explanation of the first type, with childhood nutrition as the cause that could be manipulated to affect the outcome. One could also view health as a characteristic of adult individuals that is a natural predictor of productivity.

As stressed before, there are generally not unique answers to these questions. In the case of the association between height and earnings, one researcher might find that health is the omitted attribute, and another researcher might find that childhood nutrition is the omitted

cause. Both could be right, and which answer is more useful will depend on the context. The point is that the finding that height itself is correlated with earnings is the starting point for an analysis that explores causes of earnings, that is, alternative models for earnings determination, that would reproduce the correlation between height and earnings without the implication that intervening in height would change earnings.

What does this mean for statistical practice?

A key theme in this discussion is the distinction between causal *statements* and causal *questions*. A reverse causal question does not in general have a well-defined answer, even in a setting where all possible data are made available. But this does not mean that such questions are valueless or that they fall outside the realm of statistics. A reverse question places a focus on an anomaly—an aspect of the data unlikely to be reproducible by the current (possibly implicit) model—and points toward possible directions of model improvement.

It has been (correctly) said that one of the main virtues of the potential outcome framework is that it motivates us to be explicit about interventions and outcomes in forward causal inference. Similarly, one of the main virtues of reverse causal thinking is that it motivates us to be explicit about what we consider to be problems with our model.

In terms of graphical models, the anomalies also suggest that the current model is inadequate. In combination with the data, the model suggests the presence of arrows that do not agree with our *a priori* beliefs. The implication is that one needs to build a more general model involving additional variables that would eliminate the arrow between the attribute and the outcome.

By formalizing reverse casual reasoning within the process of data analysis, we hope to make a step toward connecting our statistical reasoning to the ways that we naturally think and talk about causality. This is consistent with views such as Cartwright (2007) that causal inference in reality is more complex than is captured in any theory of inference. The basic idea expressed in this paper—that the search for causal explanation can led to new models—is hardly new (see, for example, Harris, 1999, for an example from inventory control). What we are really suggesting is a way of talking about reverse causal questions in a way that is complementary to, rather than outside of, the mainstream formalisms of statistics and econometrics.

Just as we view graphical exploratory data analysis as a form of checking models (which maybe implicit), we similarly hold that reverse causal questions arise from anomalies—aspects of our data that are not readily explained—and that the search for causal explanations is, in statistical terms, an attempt to improve our models so as to reproduce the patterns we see in the world.

References

Ashraf, Q., and Galor, O. (2013). The "out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review.*

Banerjee, A. J., and Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* New York: PublicAffairs.

Cartwright, N. (2007). *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press.

Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. New York: Norton.

Fearon, J. D. (1991). Counterfactuals and hypothesis testing in political science. *World Politics* **43**, 169–195.

Fung, K. (2013). Causal thinking. Numbers Rule Your World blog, 17 July. `http://junkcharts.typepad.com/numbersruleyourworld/2013/07/causal-thinking.html`

Gelman, A. (2011). Causality and statistical learning. *American Journal of Sociology* **117**, 955–966.

Gerber, A. (1998). Estimating the effect of campaign spending on Senate election outcomes using instrumental variables. *American Political Science Review* **92**, 401–411.

Harris, B. (1999). Pipeline inventory: the missing factor in organizational expense management. *National Productivity Review* **18**, 33–38.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*, second edition. University of Chicago Press.

Lakatos, I. (1978). *Philosophical Papers*. Cambridge University Press.

Levitt, S. D. (1994). Using repeat challengers to estimate the effect of campaign spending on election outcomes in the U.S. House. *Journal of Political Economy* **102**, 777–798.

Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason, J. P., and Pellom, A. C. (1989). Empirical Bayes procedures for stabilizing maps of U.S. cancer mortality rates. *Journal of the American Statistical Association* **84**, 637–650.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Translated and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science* **5**, 463–480 (1990).

Pearl, J. (2010). *Causality: Models, Reasoning and Inference*, second edition. Cambridge University Press.

Rubin, D. B. (1974b). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics* **2**, 1–26.

Sloman, S. A. (2005). *Causal Models: How People Think About the World and Its Alternatives*. Oxford University Press.

Yamamoto, T. (2012). Understanding the past: statistical analysis of causal attribution. *American Journal of Political Science* **56**, 237–256.