How Blindly 'Following the Data' Can Shipwreck the Best of Ideas

Published 01/22/24 10:00 AM ET

Sheldon H. Jacobson, Ph.D.

The concept "following the data" was used ubiquitously during the COVID-19 pandemic — and for good reason. With a novel virus circulating and exhibiting highly variable and unpredictable health outcomes, the only anchor that public health decision-makers had available to them was the data being collected, often in real-time. However, data in and of itself does not possess some magical and mystical power. The key is how data are collected and used.

Yet, data alone may not be reliable for predicting the future. Worthwhile analysis requires models like those available in artificial intelligence and statistics to make sense of the data and provide some understanding of what has occurred. This understanding then may be used to forecast the future, though the assumptions required to make such forecasts effectively render such forecasts highly noisy, often over a limited time horizon, and in some cases, useless.

Much like how a broken clock is right twice per day, some forecasts, albeit unreliable, may be correct. The challenge is producing accurate forecasts consistently under a wide swath of conditions.

Data is like looking in the rearview mirror. Yet, making progress requires forward motion, fraught with uncertainty. That is why models used to predict the direction of the COVID-19 virus were often consistently (and understandably) wrong.

Yet, many want to know the future. Whether it be in politics, finance, economics or sports, we listen to pundits who pontificate on trends and what lies just around the future. Sometimes they are right; other times they miss the mark, overshadowed by the few who happen to get it right.

Here are two of the areas in which "following the data" is likely to get you lost:

Election Forecasting

[Pollsters offer a plethora of forecasts of who will win elections](). As we approach Election Day in November 2024, the media swings and sways with each new poll. The [polls for who will win the 2024 presidential election]() often focus on the former president, Donald Trump, versus the incumbent president, Joe Biden, although it remains to be seen if either [will be their respective party's nominee]().

All such [forecasts are grounded in sound statistical theory](), offering informative forecasts provided that the sample of those polled resembles the people who will cast their votes on election day. The challenge is identifying such samples, since no one knows with certainty who will show up to cast a ballot.

As was seen in the 2016 presidential election, [most polls had Hillary Clinton defeating Trump in several key states](). The problem was that the sample of those polled did not align with who showed up to vote and how they voted. As a result, polls missed the mark, badly, etching a dark shadow on data in general and polling in particular.

Forecasting who will vote on election day is what makes it difficult to give credence to polls, particularly when the numbers in a race are close. Given the turbulent and polarized political environment, this makes polling data shaky for the most contested races, when it is needed the most.

Financial and Economic Forecasting

The same holds true with [financial forecasts](). At the end of each day, the direction of markets is explained based on data. Yet, much of such data are available at the beginning of the day. Cherry-picking data to explain what happened is easy; using such data to forecast the future is difficult.

[Expert opinions can also be a source of data](). Highly knowledgeable people give their view of what they see and provide informed perspectives. The challenge is that two groups of experts may come up with diametrically opposing points of view. One such group is likely to be right, while the other group must then explain why they were ill-informed, often using even more data to retrospectively analyze the missed outcome.

When it comes to economics, the [fundamental theory to which economists](#) subscribe may also explain such different perspectives and forecasts.

This is particularly true with government forecasts. The [Congressional Budget Office](#) (CBO) may provide its view of the economy, yet there are numerous factors that can change, so the accuracy of such forecasts is murky at best. The CBO's report on [the economy and budget deficits through 2033](#) is certain to be way off from what will likely occur.

Recall that the [United States last reported a budget surplus in 1998](#), with rosy pictures forecasted through 2014. We all know that such surpluses never materialized, with [sizeable budget deficits being the norm since then](#). Given that each administration has its own economic policies, changes in such directions make long-term forecasts highly unreliable.

Data is wonderful, until it is not. Their value is highly dependent on models that attempt to transform data into useful information. As the noted statistician George Box said, "[All models are wrong, some are useful](#)." The same can be said about data.

Data is important to support decision-making. Data may steer us to safety through treacherous terrain or shipwreck us on rocks. How it is collected, its source and how it is used determines its ultimate value.

[Sheldon H. Jacobson](#), Ph.D., is a professor of computer science at the University of Illinois Urbana-Champaign. A data scientist, he applies his expertise in data-driven risk-based decision-making to evaluate and inform public policy.