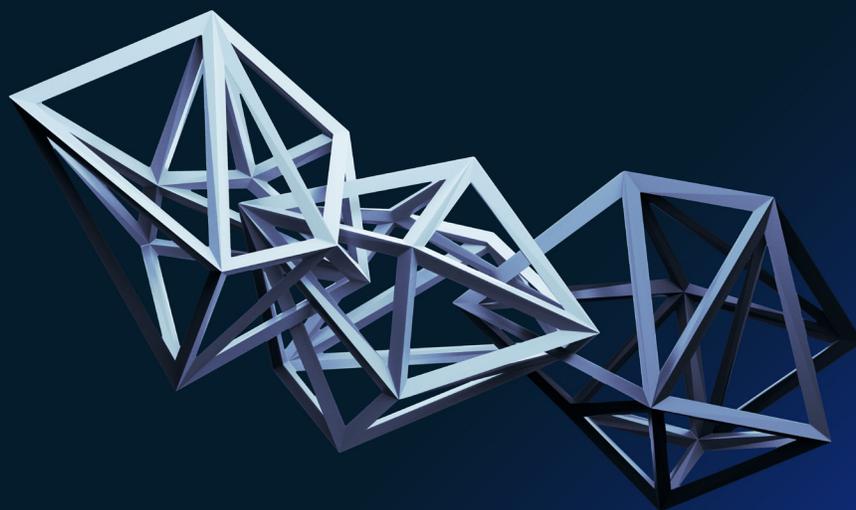


Risk Practice

Demystifying modeling: How quantitative models can—and can't— explain the world

The COVID-19 crisis has brought quantitative models to the forefront. Here are some ways that modeling helps us—as long as we avoid its pitfalls.

by Sarun Charumilind, Anas El Turabi, Patrick Finn, and Ophelia Usher



One of the many impacts of the COVID-19 crisis has been to highlight the role of quantitative models in our lives. Ideas associated with modeling, such as flattening the curve of disease transmission, are now regularly discussed in the media and among families and friends. Across the globe, we are trying to understand the numbers and what they mean for us.

Forward-looking models aren't new. They have long played an important but unseen role in day-to-day life—for instance, in pricing homeowners' insurance, anticipating the weather, and deciding how many iPhones to manufacture. However, in the COVID-19 pandemic, the scale of impact and the level of uncertainty have introduced new challenges—and notoriety—for modelers.

Used properly, models provide information that can present a framework for understanding a situation. But they aren't crystal balls that state with certainty what will happen, and they don't in themselves answer the difficult question of what to do. The eminent British statistician George Box summarized the point with his famous aphorism: "All models are wrong, but some are useful." And he refined it by saying, "Since all models are wrong, the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad."

This article explains how models can help us make sense of the world and why they behave the way they do (see sidebar "What is a model?"). We also discuss the most common modeling pitfalls and how to avoid them.

The power of models

Making decisions in the face of uncertainty is challenging, particularly during a pandemic. Quantitative models can help us understand systems and behaviors in a number of useful ways that help navigate this ambiguous environment.

Clarifying which drivers matter

Models structure data in support of reasoned decision making by restricting variables to those that matter for a particular question. For example, when developing a demographic model to help civic

leaders plan future community needs, key drivers could be birth rates, death rates, and new-job creation. Models can help users understand what is known about each element and identify the areas of continuing uncertainty.

Determining how much an input can matter

Models are well suited to exposing sensitivities: they show how even small changes in key assumptions can produce large variations in outcomes, helping decision makers establish priorities. An obvious case in point related to the COVID-19 pandemic is the massive impact of even small adjustments in the transmission rate of infection. By establishing sensitivities, models pinpoint areas for investment of effort or money to reduce uncertainty.

What is a model?

A key feature of models is that they are based on simplifications of reality. Think of a model as a map—a representation of the world with inputs that help one make decisions. A map enables map readers to make decisions about which route they might take to reach their destination. It doesn't tell them what to do nor choose their means of transportation. But making good use of a map requires an understanding of parameters such as scale and other inputs—in other words, exactly which aspects of reality have been simplified and how.

A model is a representation of a real system. Often leveraging mathematical equations, a model can test hypotheses and assumptions about a system's behavior. For example, economic models provide logical mapping of interactions within an economy, enabling economists to track the effects of a change in circumstances (such as government spending or taxes) on economic outcomes.

Facilitating discussions about the future

Models expose how different assumptions lead to different outcomes. Through discussion of modeling results, decision makers can form a collective judgement on scenarios to plan for, based on the multiple variables considered, and thus reach practical decisions (see sidebar “Building a quantitative model while using it”). For example, models were used to enable policy makers to weigh the benefits of requiring seatbelts against the moral hazard of encouraging people to drive faster. Not only do models trigger discussion, but they may force a more nuanced and evidence-based approach to decision making. In many cases, that is more important than the specific output itself.

Pitfalls to avoid when using models

A model is simply a tool, and, as with any tool, its value highly correlates with the way it is used. Models can be broken down into three main components: raw data, assumptions that define what the model does with the data, and final output. The relative importance of assumptions and data varies by model. Google Search’s autofill, for example, is mostly data driven, while the adage about waiting an hour before swimming is driven by assumptions. Each part must be viewed with a critical lens; failure to do so can lead to poorly informed decisions.

Overlooking the fact that a model can’t fix bad data

A model is only as good as its underlying data, and data in a time of extreme uncertainty, such as a global pandemic, present a serious challenge. Just as rotten ingredients won’t produce a tasty dish no matter how good the recipe is, poor data lead to poor output from a model.

Data can be wanting for various reasons: too few data points, inconsistency, inaccuracy, or incorrectly generalizing from a particular data set. Modeling anything related to a novel virus entails the risk of using bad data. Virtually all the data series being collected about the COVID-19 crisis are incomplete or subject to caveats. For example, using data on the impacts of the COVID-19 pandemic in one

Building a quantitative model while using it

Extreme uncertainty is one of the most pronounced and disturbing characteristics of the COVID-19 crisis. That is what makes models potentially useful in the effort to understand the virus. But it also makes the very business of constructing and validating models uniquely challenging.

Most models in general use, such as economic models that underpin business transactions, are known quantities. The data and assumptions fed into them are well understood, and they have benefited from a process of continuous improvement and refinement over many years.

However, in a pandemic caused by a novel virus, uncertainty pervades almost every aspect, from the basics about the biological behavior of the virus to the reliability and consistency of the data being collected to the way that public-health interventions could influence infection rates. That poses a modeling challenge akin to flying a plane whose key components and structures are still under construction and require continual revisions.

geography to model potential impacts in another community can be problematic. Data might not be generalizable if the populations differ in important dimensions, such as age.

Taking assumptions and simplifications for granted

Assumptions aren’t facts; they should be subject to regular, searching review (see sidebar “The risks of bias in modeling”). For example, prior to the

A model is only as good as its underlying data, and data in a time of extreme uncertainty, such as a global pandemic, present a serious challenge.

2008 financial crisis, a key assumption in multiple models was that real-estate prices wouldn't see major declines. Values had consistently increased in the precrisis years, so some began to take that assumption as a fact, thereby obscuring other possible scenarios.

Assumptions aren't static; they are subject to change as we learn more, especially in novel circumstances. Estimated rates of death from COVID-19 have been constantly revised as our understanding has expanded. Models tell you what might happen if you believe specified things about different variables. Those ifs all need to be revisited frequently if the model is to remain relevant and useful.

Expecting too much certainty

Models aren't designed to eliminate uncertainty but to limit the range of uncertainties in a given situation by showing what might happen in a variety of defined scenarios. Uncertainty can arise from the very structure of the model, basic assumptions, and ongoing data inputs. For example, hurricane models are an attempt to gain understanding of where hurricanes might make landfall. The models start with significant uncertainties around the path the hurricane might take, and the uncertainties decrease over time as landfall nears.

Usually, models provide guidance on possible futures given multiple inputs (see sidebar "Modeling philosophy for the COVID-19 pandemic"). That makes it dangerous to take a subset of a model's outputs at a particular point in time as a singular reality. For example, in addition to a popular model for tracking COVID-19-related deaths and

The risks of bias in modeling

A model, the data underlying a model, and a model's use can all be subject to biases. Intentionally or unintentionally, those factors can amplify or validate biases. Biased data can lead to faulty decisions, while a user's preconceptions can lead to an incorrect conclusion about model outputs.

Bias often affects the data-collection phase. An example is when COVID-19-testing results in a given territory are assumed to provide a complete picture of prevalence but don't, because certain communities haven't been adequately tested. Without correcting for such biases, data that purport to reflect the entire population of a territory are distorted.

hospital demand, the Institute for Health Metrics and Evaluation has released a model that predicts daily infections and testing. For August 2, 2020, it predicts 80,130 infections, which seems very precise (and quotable). However, closer inspection appropriately shows a range of 45,595 to 156,889 infections.¹ That is a huge range, but it doesn't negate the usefulness of the model. It is an important indicator of the level of uncertainty that should be taken into account when making any subsequent decisions.

¹ "COVID-19 projections," Institute for Health Metrics and Evaluation, June 15, 2020, covidhealthdata.org.

Ultimately, when using models to make decisions or when interpreting their outputs, there are several key questions to ask: How has this model simplified the world? What inputs does the model require, and how knowable, certain, and stable are those inputs? What are the outputs telling us, and what is the level of uncertainty? And lastly, how have users engaged with this model in the process of making decisions?

Modeling philosophy for the COVID-19 pandemic

McKinsey uses models in a variety of ways. During the COVID-19 pandemic, we have created models to develop scenarios with the aim of helping decision makers understand the evolution of the situation and the potential order of magnitude of its impacts. Such understanding can support decision makers' discussions on what actions to take.

Continually refined with the latest data, the scenarios aren't an attempt to predict the future, nor are they an attempt to compete with peer-reviewed simulations published by academic research groups. Rather, they are designed to create some range around the uncertainties inherent in the situation, provide clarity about the criticality of different assumptions in driving outputs, and support discussion around first- and second-order implications.

Satisfactory answers to these questions will foster a better understanding of potential future scenarios and better decisions in an evolving and uncertain situation. (For a list of suggested reading about the use of models in the current crisis, see sidebar "Read me: Quick hit on COVID-19 models.")

Read me: Quick hit on COVID-19 models

To learn more about modeling related to the COVID-19 crisis, consider reading the following publications:

- Ryan Best and Jay Boice, "Where the latest COVID-19 models think we're headed—and why they disagree," *FiveThirtyEight*, June 24, 2020, projects.fivethirtyeight.com
- Caroline Buckee and Inga Holmdahl, "Wrong but useful—what Covid-19 epidemiologic models can and cannot tell us," *New England Journal of Medicine*, May 15, 2020, nejm.org
- Peter H. Gleick, "No COVID-19 models are perfect, but some are useful," *TIME*, May 19, 2020, time.com
- "COVID-19 estimation updates," Institute for Health Metrics and Evaluation, June 2020, healthdata.org

Sarun Charumilind is a partner in McKinsey's Philadelphia office; **Anas El Turabi** is an expert in the Waltham, MA, office; **Patrick Finn** is a senior partner in the Detroit office; and **Ophelia Usher** is an expert in the New York office.

The authors wish to thank Jessica Lamb, Sean McCormick, and Aspen Wang for their contributions to this article.

Designed by Global Editorial Services
Copyright © 2020 McKinsey & Company. All rights reserved.