

Measuring and Managing COVID-19 Model Risk

Mark J. Jensen

Federal Reserve Bank of Atlanta

Summary:

One of the many lessons learned from the financial crisis is the increased awareness of model risk. In this article, I apply the best practices of model risk management found in SR 11-7 (which offers regulatory guidance on the best practices for managing model risk) to COVID-19 models. In particular, I investigate the Institute of Health Metrics and Evaluation's (IHME) model to see if it has been effectively challenged with a critical assessment of its conceptual soundness, ongoing monitoring, and outcomes analysis.

Key findings:

1. Open source COVID-19 models and public data lend themselves to independent and well-informed model validation.
2. Effective challenge of the IHME model has improved it and is helping to inform key stakeholders of the model's intended use and limitations.

JEL classification: C1, C11, C52

Key words: COVID-19, model risk management, SR 11-7

<https://doi.org/10.29338/ph2020-07>

**CENTER FOR QUANTITATIVE
ECONOMIC RESEARCH**

**CENTER FOR HUMAN
CAPITAL STUDIES**

**CENTER FOR FINANCIAL
INNOVATION AND STABILITY**

**CENTER FOR HOUSING
AND POLICY**



The Federal Reserve Bank of Atlanta's Policy Hub

leverages the expertise of Atlanta Fed economists and researchers to address issues of broad policy interest. Our research centers coordinate this work and seek to influence policy discussions. Areas of interest include: forecasting, fiscal policy, and macroeconomics (Center for Quantitative Economic Research); financial stability, innovation, and regulation (Center for Financial Innovation and Stability); human capital, labor markets, health, and education (Center for Human Capital Studies); and government-sponsored entity reform, mortgage markets, and affordable housing (Center for Housing and Policy). Sign up for email updates at frbatlanta.org/research/publications/policy-hub.

Measuring and Managing COVID-19 Model Risk

Summary: One of the many lessons learned from the financial crisis is the increased awareness of model risk. In this article, I apply the best practices of model risk management found in SR 11-7 (which offers regulatory guidance on the best practices for managing model risk) to COVID-19 models. In particular, I investigate the Institute of Health Metrics and Evaluation's (IHME) model to see if it has been effectively challenged with a critical assessment of its conceptual soundness, ongoing monitoring, and outcomes analysis.

JEL Classification: C1, C11, C52

Key words: COVID-19, model risk management, SR 11-7

About the Author:

Mark J. Jensen is a vice president, a senior economist, and the team leader of the financial markets team in the Research Department of the Federal Reserve Bank of Atlanta. Dr. Jensen concentrates his research on Bayesian financial econometrics. During the Financial Crisis, he was heavily involved with the Federal Reserve System's Comprehensive Capital Analysis and Review by participating on its Risk Evaluation Team for retail products and by validating the Federal Reserve's asset-backed security models for the Model Validation Group.

Acknowledgments: The author thanks Neil Desai, Mark Fisher, and Paula Tkac for their helpful comments and suggestions. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Comments to the authors are welcome at Mark.Jensen@atl.frb.org.

Policymakers, healthcare officials, business owners, the media, and the public are looking for any similarities between the financial crisis of 2008 and the current COVID-19 pandemic that could help them in making better and more informed decisions. Though the underlying events of the two crises are very different, there are, however, lessons learned from the financial crisis that also apply to the pandemic. A common theme playing out in the pandemic—one that was present before, during, and after the financial crisis—is the modeling and forecasting of hard-to-predict, unknown values. For example, in the case of the financial crisis predicting the number of future mortgage defaults on a bank's portfolio was unpredictable. For the virus, predicting the number of future COVID-19 related fatalities for a particular geographical area is challenging.

At the core of the financial crisis was the challenge of modeling the value of structured credit products like collateral debt obligations (CDO), which the rating agencies used to assess the credit risk of these products.¹ In a pandemic crisis, these challenges include the modeling of the number of future fatalities, the rate of infection, the future level of hospital resources, the likelihood of a second wave, the other unknown values associated with the COVID-19 virus, and how the public might respond to changes in social distancing. Like the credit rating agencies, having a model to forecast these unknown values related to COVID-19 help key stakeholders assess the risks and navigate them appropriately.

To make accurate predictions, we need good models, but modeling is hard.² Modeling is even harder when data are sparse, occasionally mismeasured, or unavailable or—like the credit rating agencies' experience with CDOs—no data are available through a credit or business cycle. Similarly, in the case of COVID-19, modelers do not have the luxury of observing the daily number of fatalities or the rate of infection from the virus's beginnings to its eradication.

Model Risk Management

Good modeling is just as much a subjective art form as it is an objective science. In addition to requiring high levels of technical and mathematical skills, modeling involves learning by doing, knowing the institutional details, understanding the nuances of the data, recognizing the empirical regularities, etc. Those who already have experience modeling pandemics are most likely more capable of proposing the initial models. However, feedback, input, and critiquing from others are critical to improving these initial

¹ See Charles Smithson, "Valuing hard-to-value assets," *Risk Magazine*, September 2008.

² See Maggie Koerth, Laura Bronner, and Jasmine Mithani, "Why It's So Freaking Hard to Make a Good COVID-19 Model," *FiveThirtyEight*, March 31, 2020 <<https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a-good-COVID-19-model/>> (accessed June 16, 2020).

models. This iterative process—in other words, the practice of model risk management—is one of the primary lessons banks and their regulators learned from the financial crisis, and it also applies to the modeling of COVID-19.

As a consequence of financial institutions relying on incorrect models or using them in an unintended manner, the Board of Governors of the Federal Reserve System and the Office of the Comptroller of the Currency jointly issued guidance to banking organizations and supervisors in the form of SR 11-7 on the best practices for managing model risk.³ Model risk is the potential for adverse consequences from decision makers (for instance, policymakers, healthcare officials, and business leaders) making decisions based on incorrect or misused model output or conflating model prediction with certainty. In the case of COVID-19, model risk can lead to an increase in human loss as well as financial and economic ruin as a result of poor decision making based on inappropriately used or misunderstood model results.

SR 11-7 has fundamentally improved how model developers, chief risk officers, and even the boards of directors of financial institutions manage model risk. SR 11-7 also led the supervising agencies to increase their human capital in the area of quantitative modeling skill to better monitor a financial institution's level of model risk. Through an ongoing, coordinated effort on managing model risk between supervisors and financial institutions, the capital planning processes of financial firms are more robust, and banks are now more resilient to adverse shocks like COVID-19. How well, then, do the COVID-19 models—and in particular, the Institute for Health Metrics and Evaluation's (IHME) model of fatalities related to COVID-19⁴—stand up to the best practices found in SR 11-7?

At the heart of SR 11-7 is “effective challenge” – the critical analysis by objective, informed parties that can identify model limitations and produce appropriate changes. Effective challenge consists of three areas: conceptual soundness, ongoing monitoring, and outcomes analysis. Because financial firms use proprietary data to develop their models, effective challenge is formally performed by an in-house model validation team that is independent of the model developers. COVID-19 models and their

³ See SR Letter 11-7 and OCC 2011-2, “Guidance on Model Risk Management,” April 4, 2011 <<https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>> (accessed June 16, 2020).

⁴ See IHME COVID-19 health service utilization forecasting team, “Forecasting the Impact of the First Wave of the COVID-19 Pandemic on Hospital Demand and Deaths for the USA and European Economic Area Countries,” <<https://doi.org/10.1101/2020.04.21.20074732v1>> (accessed June 16, 2020).

data⁵ are mostly in the public domain.⁶ Hence, the process of effective challenge can be done informally and effectively through open comment and critical peer review⁷, not the closed process of blind journal refereeing.

Conceptual Soundness

Conceptual soundness involves the design, theory, and purpose of the model. To help healthcare officials plan for the demand for medical services, model developers of the original IHME model⁸ fit a statistical model to the cumulative number of COVID-19-related fatalities as a function of social distancing.⁹ It was not an epidemiological model capable of forecasting the long-run behavior of the virus. Instead, the modelers assumed daily fatalities grow exponentially until reaching their peak level and then extrapolated forward from that curve. Given this modeling assumption, the original IHME model should not be used to infer the number of daily fatalities beyond seven to 21 days, nor should it be used once the number of fatalities in a location has peaked.

Sound model risk management, as spelled out in SR 11-7, explicitly states that model users should only use a model for its intended purpose. Since the original IHME model predicted how the introduction of social distancing measures would affect fatalities, and not how fatalities would be affected when such policies are relaxed, policymakers, healthcare officials, and the media should not have relied on the model to indicate when social distancing measures could safely be relaxed. Instead, the original IHME model was useful to measure how social distancing was helping to flatten the curve.

⁵ For data on confirmed cases and deaths see “Data on COVID-19 (coronavirus) by *Our World in Data*” <<https://github.com/owid/covid-19-data/blob/master/public/data/README.md>> (accessed June 16, 2020). Google is making available data on human mobility, to quantify the changing impact social distancing has on deaths, at “COVID-19 Community Mobility Reports” <<https://www.google.com/covid19/mobility/>> (accessed June 16, 2020).

⁶ For an ensemble of models forecasting the number of national- and state-level deaths, see the Centers for Disease Control and Prevention website <<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html>> (accessed June 16, 2020).

⁷ For instance, see the critical assessment of the IHME model by Nicholas Jewell, Joseph Lewnard, and Britta Jewell, “Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic,” *Annals of Internal Medicine*, April 14, 2020 <<https://doi.org/10.7326/M20-1565>> (accessed June 16, 2020).

⁸ In response to additional data and an increasing understanding about the behavior of the virus, the IHME expanded its model to a hybrid, multistage approach that now forecasts cumulative deaths in response to testing and increasing social interaction. See “COVID-19: What’s New for May 4, 2020,” COVID-19 Estimation Updates, Institute for Health Metrics and Evaluations,” <<http://www.healthdata.org/COVID/updates>> (accessed June 16, 2020).

⁹ See IHME COVID-19 health service utilization forecasting team, “Forecasting the Impact of the First Wave of the COVID-19 Pandemic on Hospital Demand and Deaths for the USA and European Economic Area Countries,” <<https://doi.org/10.1101/2020.04.21.20074732v1>> (accessed June 16, 2020).

The potential economic and social costs from incorrectly using a model should encourage all stakeholders to have a sound understanding of model risk by knowing the correct use of a model as well as its limitations.

Conceptual soundness also includes ensuring data quality and relevance. One reason CDOs were challenging to model before and during the financial crisis was that there were no reliable, granular-level data on the assets underlying these structured credit products.¹⁰ Even for the most sophisticated models, garbage data in lead to garbage model predictions out.

Minimizing model risk means it's important to point out the limitations and potential shortcomings of one's data. In the case of COVID-19, there was initially only sparse data on how fatalities from the virus would grow, how long it took for fatalities to peak, what level they would peak at, and how daily fatalities would behave after peaking. For example, for states that experienced fatalities, some of them had only a few days' worth of data.¹¹

For some time, designing models for sparse and noisy data has commanded the attention of Bayesian statisticians and econometricians.¹² Bayesians deal with limited data by learning and borrowing from others who have already experienced a similar event and collected data from that event. Learning and borrowing information from others with similar experiences is reflected in the construction of a set of initial beliefs called the Bayesian prior.¹³ The prior captures what the modeler initially expects to occur for the location where data are limited. As the modeler collects and analyzes data from a particular location, the modeler updates the prior's initial beliefs.

In the IHME model, the model developers initially used proxy data from Wuhan City, China, to form the prior for how the peak time responds to social distancing. Early on, Wuhan was the only location where fatalities had peaked and had also enacted social distancing, so the modelers borrowed and learned from these data. Over time, the IHME model developers added data from other locations

¹⁰ See "Shareholder Report on UBS's Write-Downs," April 2008.

¹¹ Because testing has not been random, measures of the number of infected are even noisier than reported deaths. Infection data are also prone to underreporting since testing is limited.

¹² See Mark J. Jensen, "Stress Testing with the Help of Bayes' Theorem," *Notes from the Vault*, February 2016 <<https://www.frbatlanta.org/cenfig/publications/notesfromthevault/1602.aspx>> (accessed June 16, 2020).

¹³ For a layperson's introduction to Bayesian statistics, see S.B. McGrayne, *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant From Two Centuries of Controversy*, New Haven, Yale University Press, 2011.

that had peaked in the number of daily fatalities and used this cross-section of data to update their Bayesian prior.¹⁴

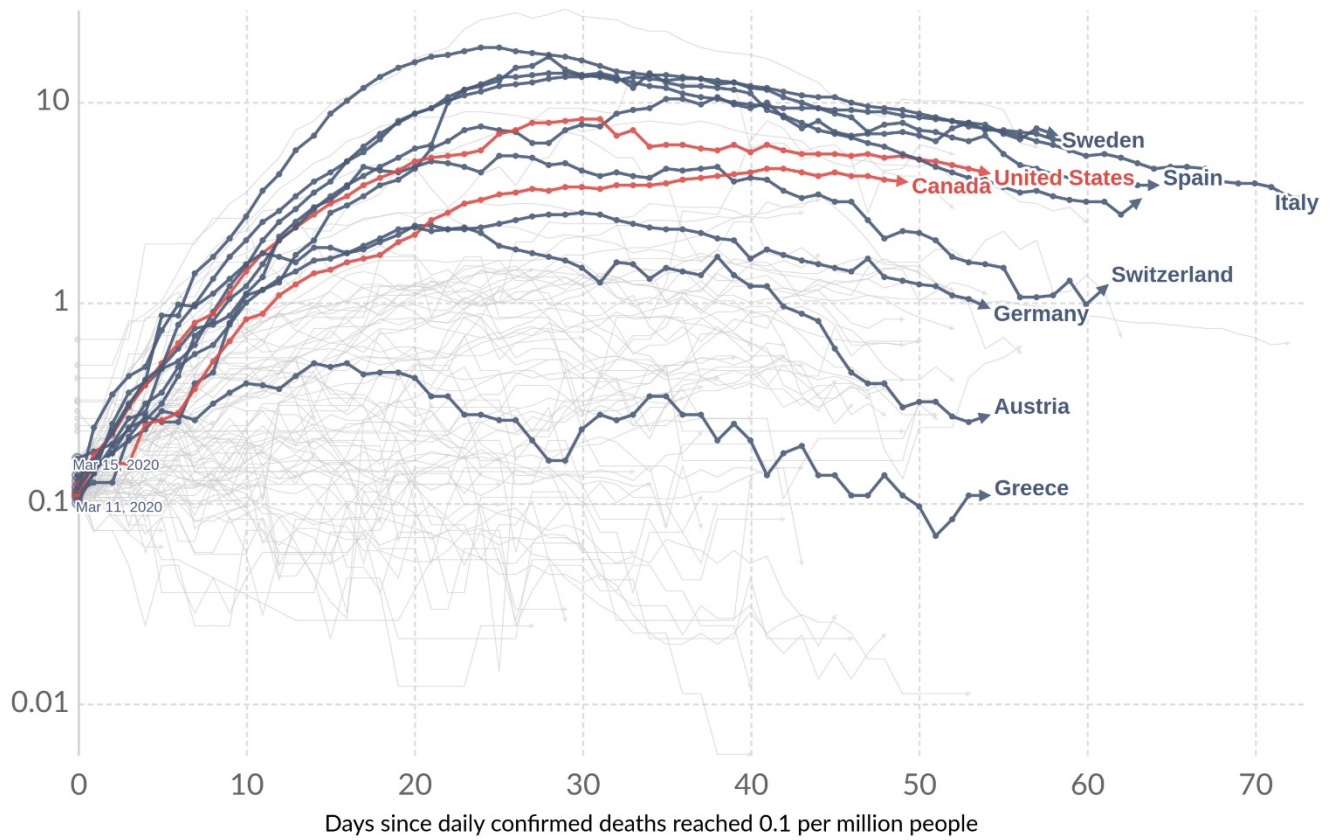
According to SR 11-7, when modelers use proxy data related to a financial institution, they should justify the data and show that they represent the financial institution's portfolio. When the relationship between social distancing and every location's peak in fatalities is treated the same, projected fatalities with the IHME model for each location will tend to be biased toward the global overall relationship. As a result, there is the potential for locations where social distancing did not affect the peak in fatalities to look like they did, and for locations where social distancing had a very large effect on their peak to look like it did not.¹⁵

In figure 1, we plot the trajectory of daily fatalities per million people for eight different countries that had already reached their peak. As the wide variety of patterns in these trajectories makes evident, COVID-19 fatalities do not follow the same dynamic path from country to country. Given the heterogeneity of these trajectories, the model developers of the IHME model should point out the potential bias of their predictions. For example, using the countries in figure 1 as an example, this potential bias means that, going forward, Sweden and Greece would behave more like Germany. Identifying such drawbacks of one's modeling approach will help policymakers and healthcare officials understand why actual fatalities in a location where data are sparse can differ dramatically from the model's predictions.

¹⁴ At the time of this writing, the IHME model used 13 locations where peak deaths had occurred to form a prior for the relationship between social distancing and peak time. A larger cross-section of such locations also helps reduce the noise and underreporting bias prevalent in the recorded number of deaths.

¹⁵ See Mark Fisher and Mark J. Jensen, "Bayesian Nonparametric Learning of How Skill is Distributed across the Mutual Fund Industry," *Journal of Econometrics*, forthcoming. They show how highly (un)skilled fund managers look average when the prior for skill is computed in a manner similar to the IHME model's prior.

Figure 1: Daily Confirmed COVID-19 Deaths per Million



Note: Data are depicted in log-scale as of May 14, 2020, using a seven-day moving average to smooth out day-to-day fluctuations.
Source: Our World in Data

Ongoing Monitoring

Ongoing monitoring involves constantly revising, improving, and possibly replacing a model in response to new information or a model not performing as intended. For COVID-19, dynamic modeling is the response to the virus's fast-moving threats and developments, along with incorporating the flood of new information. How modelers respond to this new information is where ongoing monitoring plays a critical role in effectively challenging a model.

A consequential example of not performing ongoing monitoring is the U.S. Office of Federal Housing Enterprise Oversight risk-based capital model for Fannie Mae and Freddie Mac. In the run-up to the financial crisis, these models failed to detect the increasing risk in the mortgage market because the models were estimated using stale data from 1979 to 1997.¹⁶

Ongoing monitoring has been occurring for the IHME model through critical peer review and professional feedback. By the end of May, at least five revisions to the model had been made. One notable example was the early revision to the model's conservative range of possible outcomes for future fatalities. During the period of March to April 2020, the range of uncertainty in the model's predictions failed to properly include the correct proportion of observed fatalities.¹⁷

The IHME model had been overly optimistic in its range of possible outcomes because the model's statistical properties were based on the incorrect assumption that the cumulative number of fatalities are independent from day to day.¹⁸ Also, the model developers had assumed large sample statistical behavior for the parameter estimates for locations where there were too few observations for these asymptotic properties to hold.¹⁹ The model developers promptly addressed these statistical concerns by changing the way the range of possible model outcomes were computed from an

¹⁶ See W. Scott Frame, Kris Gerardi, and Paul S. Willen, "The Failure of Supervisory Stress Testing: Fannie Mae, Freddie Mac, and OFHEO," Federal Reserve Bank of Atlanta Working Paper 2015-3, <<https://www.frbatlanta.org/research/publications/wp/2015/03.aspx>> (accessed June 16, 2020).

¹⁷ See Nicholas Jewell, Joseph Lewnard, and Britta Jewell, "Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic," *Annals of Internal Medicine* 2020, <<https://doi.org/10.7326/M20-1565>> (accessed June 16, 2020).

¹⁸ It is the daily number of fatalities that are independent, not the cumulative number.

¹⁹ See Spencer Woody, Mauricio Tec, Maytal Dahan, Kelly Gaither, Michael Lachmann, Spencer Fox, Lauren Meyers, and James Scott, "Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones," University of Texas at Austin COVID-19 Modeling Consortium, April 17, 2020, http://www.tacc.utexas.edu/ut_covid-19_mortality_forecasting_model_report> (accessed June 16, 2020).

approximate model-based uncertainty approach to a simulation-based approach.²⁰ Developers of the IHME model will need to continue addressing issues raised by their peers and the profession at a frequency appropriate for the flow of new information to avoid a stale and irrelevant model.

Outcomes Analysis

The last step to effective challenge is performing outcomes analysis. Of the many quantitative and qualitative approaches to conducting outcomes analysis, back-testing is the form of outcomes analysis most often required by internal risk governance policy. Back-testing a model involves estimating the model over a restricted range of data and using the estimated model to predict the out-of-sample data. For example, in the case of COVID-19 one would use the daily data on the number of fatalities up to a selected date, but before the end of the data set, to estimate the model. Then the estimated model could be used to forecast the number of fatalities over the data not used to estimate the model.

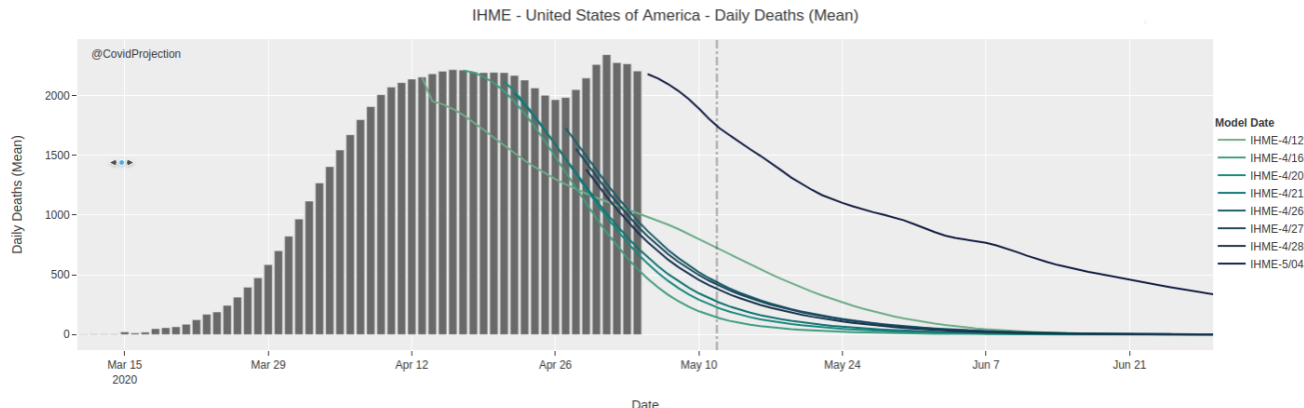
The COVID Projections Tracker (CPT) website (at <https://www.covid-projections.com/>) plots the back-testing results for the IHME model. In figure 2, we reproduce the CPT graph of the IHME model's back-testing forecasts of the daily number of fatalities for the United States for eight different in-sample dates (lines) against the actual number of fatalities (vertical bars).²¹ Clearly, from the point forecasts of figure 2, the IHME model was overly optimistic about the time to the peak and the maximum number of daily fatalities. Neither more data nor updating and revising the IHME model fixed this downward bias in the number of future daily fatalities.²² Given the common theme found in these back-testing results, policymakers and healthcare officials should have been careful and applied conservative and qualitative adjustments to the IHME model's forecasts when making decisions and informing the public about the future path of COVID-19 fatalities.

²⁰ See "Forecasting the Impact of the First Wave of the COVID-19 Pandemic on Hospital Demand and Deaths for the USA and European Economic Area Countries," sections 9.1 and 9.2 of Appendix B: CurveFit Tool and Analyses of IHME COVID-19 health service utilization forecasting team, <<https://doi.org/10.1101/2020.04.21.20074732v1>> (accessed June 16, 2020).

²⁰ Ideally, one would also like to see the model's range of uncertainty around its point forecast.

²² The considerable shift in the projected number of deaths associated with the back-test date May 4, 2020, is computed with a new multilevel IHME model that layers an epidemiology SEIR model onto the curve-fitting approach. See the IHME's COVID-19 health service utilization forecasting team, "COVID-19: What's New for May 4, 2020," May 4, 2020 <http://www.healthdata.org/sites/default/files/files/Projects/COVID/Estimation_update_050420.pdf>.

Figure 2: Back-Testing Results for the IHME Model



Source: COVID Projections Tracker

Sensitivity Analysis

Outcomes analysis of COVID-19 also includes modeling how fatalities will respond to changes in social distancing policies or to people’s willingness to adhere to these policies. This requires modelers to make explicit assumptions about what the policy will be—or how people will behave—during the period being predicted. In the case of the Federal Reserve Board’s comprehensive capital analysis and review, a stress test of bank capital, modelers predict losses under different adverse economic scenarios.²³ However, assumptions about the future are often wrong.

Model users must understand that the model’s predictions depend on the model developers’ assumptions about the future level of social distancing.²⁴ Generating multiple future possible scenarios for social distancing can help show a model’s range of possible outcomes and demonstrate the level of uncertainty in the predictions. In SR 11-7, this multiple-scenario exercise is a way of carrying out the model’s “sensitivity analysis,” which is critical to validating a model used to make predictions for conditions not seen before.

All models—be they economic, statistical, or epidemiological—are imperfect abstractions of the real world. Understanding the degree of these imperfections is at the core of model risk management. As I have pointed out, the guidance and best-practices found in SR 11-7 help minimize model risk and make good models better, and the IHME model has seen informal, peer-to-peer, and effective challenge.

²³ See “Comprehensive Capital Analysis and Review 2020 Summary Instructions,” Board of Governors of the Federal Reserve System, March 2020 <<https://www.federalreserve.gov/publications/comprehensive-capital-analysis-and-review-summary-instructions-2020.htm>> (accessed June 16, 2020).

²⁴ For example, the IHME model treats each location independently. Hence, the model explicitly does not incorporate assumptions about how an increase in future travel will affect the number of deaths.

This form of model risk management has helped the developers of the IHME model revise and redesign their model a number of times, allowing it to go from a good model to a better one. Any model forecasting the future path of fatalities related to COVID-19 should likewise be effectively challenged through independent and qualified reviewers before policymakers or healthcare officials use that model to make critical decisions.