Article from
The Modeling Platform
August 2020

SOCIETY OF ACTUARIES
MODELING SECTION

THE MODELING PLATFORM
AUGUST 2020

Reflections on an Epidemic

By Paul Conlin

On New Year's Eve 2019, my family went out to dinner with another family. Our twin boys are friends with, but go to a different school than, their son, so we took advantage of the holiday to catch up with them and spend time at one of the Chicago area's most famous, and busiest, highest-variety seafood restaurants (Bob Chinn's in Wheeling, Ill., if you're familiar). We had to wait in line for quite some time for a table for seven, but I didn't even mind—I was so grateful for the prosperous decade just past, and so looking forward to ringing in the New Year and the new decade. The 2020s, wow! Where have the years gone? I don't know about you, but I remember New Year's 2000 almost like it was yesterday. And now we were heading into the twenties! What did the future hold? Finally, the rough edges would be smoothed off of technology, the gig economy, internet trolling, vaping by minors, opioid-based painkiller dispensing—and it would be onward and upward to our bright, shining future. Right? Right?

Well, as we all know, it took all of 60 days for that hopeful plan to go off course. A pandemic. Really, people, a pandemic? With 150,000 or more dead Americans? With social distancing? And 14 percent unemployment? And second quarter 2020 GDP growth headed to minus 30 percent or minus 40 percent? And untold intangible damage to the economy for an unknowable period of time? Really??

But ... although the timing was unknowable to all of us happy diners at Bob Chinn's, the event itself—at least the medical part—wasn't, or shouldn't have been. News coverage describes it as "unprecedented." And viewed through the lens of "during our lifetime," OK, I'll give you that one. This is the worst thing that's happened in my 54-year lifetime. Yet...

ACTUARIES SHOULD TAKE A LONGER VIEW

We're actuaries. We think in longer than 54-year time spans. And through that lens? Well, this has happened before, and you don't have to go back that far to find it.

Ancient Athens had a well-documented plague occur during the Peloponnesian War. About 500 years later, during the Roman Empire of the Antonines, returning soldiers brought back smallpox with them, probably from Mesopotamia, which became endemic in Europe and was brought by their descendants 1,400 years later across the Atlantic, exposing the nonimmune Native Americans (thus killing over half of them), and was only eradicated in the 20th century. Seventy years after being laid low by the Antonine plague, the even more destructive Plague of Cyprian (so named for a Carthaginian bishop who wrote about it) ran amok through the empire from A.D. 249 to 261 or so. Malaria had long been rampant in swampy areas of Rome itself (and other swampy areas of the world where mosquitoes and humans share living space). A mysteriously thinly documented Plague of Justinian traveled through the Byzantine Empire in 541 and 542, including the capital, Constantinople, itself; paving the way for, among other things, the Slavic peoples to migrate into central and eastern Europe. The bubonic plague made use of the roads that Genghis Khan's heirs opened and maintained through central Asia to travel on fleas, which traveled on rats, which traveled with merchants and sailors (it arrived in Europe in 1347 when a merchant ship made port at Messina, Sicily), and infected and killed possibly one-third of all humanity from 1347 to the mid-1350s.

Travel. Note that theme.

The otherwise-prosperous and progressive 19th century struck an unfortunate balance of intermingling peoples and nations nearly as much as the already lamented, dearly departed 2010s, but without having the medical understanding of the 20th and 21st centuries. And the end result was a toxic brew of pandemics: cholera, polio, typhus, diphtheria, yellow fever, and, yes, influenza. The 1918–19 Spanish Flu spared almost no one, sickening even King George V of Britain and U.S. President Woodrow Wilson. Unlike COVID-19, it was particularly harsh on the young and healthy and contributed to a deep but brief depression in 1920–1921, which could be preview of the world economy in 2020–2021.

So, no, the COVID-19 pandemic is not unprecedented. Its likes have happened before, and will happen again. But enough history. You're here to read about modeling. Let's get to it.

MODELING AN EPIDEMIC USING DIFFERENCES-IN-DIFFERENCES

The cholera epidemic was a scary one. Even though cholera (which had been locally described by Portuguese and British merchants and likely was already endemic on the Indian subcontinent in the 1700s) broke out around the world in 1817, as late as 1855 its origin (was it bacteria? a fungus? "bad air"?) was still unknown. The man who solved it was one of the founders of epidemiology, John Snow.

Dr. Snow was a physician, but he was also a data modeler, as skilled in modeling as any human living in 2020. [Victorian Britain gave us a dream team of modelers and statisticians: Sir Francis Galton (1822–1911) invented the word "regression" as we understand it today in its regression-to-the-mean sense, studying the heights of successive generations of family members.] In an era without computing, Snow relentlessly collected data on deaths by subdistrict of Britain in different outbreaks (one in 1849, one in 1854) and compared the water companies of each district. This was complicated by two different companies sometimes having unique service areas, and then sometimes both serving the same towns.

Dr. Snow had a hunch, which turned out to be correct, that one of the companies was transmitting a germ (which turned out to be the bacteria that causes cholera, which reproduces and grows in just about any standing water it can embed itself into) via its water pump. He invented a statistical methodology, still used by economists today, called differences-in-differences. (The differences in Snow's model were the Deltas of the death counts in water districts using company A versus the death counts in water districts using company B versus the death counts in districts using companies A and B; the second dimension of differences was deaths in the 1849 outbreak versus in the 1854 outbreak.[1])

Differences-in-differences is a twist on multiple linear regression, in which the coefficients include the usual *x*-intercept (alpha), two coefficients (beta and gamma) of the independent variables, but then a differences-in-differences "dummy variable" with the coefficient (lowercase delta). Given enough data, the four regression coefficients can be solved for and divided by their standard deviation to get a *t*-statistic and statistical significance. If the delta coefficient is statistically significant, its value can be assumed to be the quantification of the value of the interaction of the independent variables.

This is the part of the article where I should start showing formulas. But I'm not going to do that. Because with differences-in-differences, there is no one hypothesis formula. It completely depends on the number of independent variables and the combinations of those independent variables that are assumed to interact versus those that are not. The hypothesis is completely at the discretion of the modeler.

Economists (most prominently Alan Krueger—sadly, just deceased in 2019) have used differences-in-differences for all kinds of experiments, often testing the economic impact on market participants from various forms of regulation or other underlying exogenous or endogenous factors. What the successful models do appear to have in common is that the underlying is trending; and various counterfactual trends can be examined via differences-in-differences models. The trending property of the underlying data makes it particularly interesting to actuaries.

Emanuel Derman, formerly of Goldman Sachs and Columbia University, once gave a good analogy for imagining the equation first, then testing it. One of the most important, and strangest, laws of physics is Kepler's Second Law, which states that "the line between the sun and a moving planet sweeps out equal areas in equal times" (e.g., a planet like Mercury, which is close to the sun, orbits very quickly).[2] How did Kepler, using 16th-century astronomical technology, discover it? Notice the first two words in the quote above: "the line" between the sun and a planet. Do you see the insight? The insight is this: There is no actual line between any two objects in outer space. Kepler completely imagined one, then tested out the results against his observations. Only upon seeing that the calculations worked did he conclude he had stumbled upon a law.

Hypothesis/model first, calculations/validation second.

Brilliant.

A DIFFERENCE IN PERSPECTIVE: FINANCE VS. MATHEMATICS

I conclude with an actuarial example. In 2019, *North American Actuarial Journal* published an article on property and casualty incurred but not reported (IBNR) runout from 1994 to 2010, written by three finance professors.[3] The authors downloaded Schedule P data from all of the orange blanks in the National Association of Insurance Commissioners database and did a multivariable linear regression against the 20 or so characteristics of the companies for all of the 26 or so lines of business that P&C companies write.

They tested the statistical significance of the coefficients and were fortunate to find a nice dispersion of a few, but just a few, significant coefficients at the 90th, 95th and 99th percentiles. (My personal favorite "discovery" for health actuaries was that medical malpractice coverage has had favorably developing reserves for the past quarter century, which has contributed to favorable medical trend for health plans. The asterisk indicates 90 percent statistical significance.)

But note the difference in the economics or finance approach versus the mathematical approach. The finance professors viewed the coefficients as the repositories of data for analysis and conclusions, not the dependent variable.

Come up with your equation first. Then see if it works.

WHERE DO THE MATHEMATICIANS GO WRONG?

Mathematicians and probabilist statisticians have been plowing ahead during the past half century in this field. An iconoclastic Russian mathematician named Kolmogorov worked out the underlying field paradigm of probability events in the 1930s, which led to the development of stochastic processes. This finding led to ergodic theory, which is the study of processes that repeat in cycles, and what's happening "inside the system" upon cyclical repetition of the process: Is the original starting state replicated (or nearly replicated), or is it permanently unrecoverable (i.e., has "mixing" occurred)?

A wonderful way to learn theoretical math for actuaries who don't want to get bogged down in proofs that one plus one equals two, but who still respect the insights and structure offered by pure mathematics, is counterexamples. Counterexamples were originally developed to test the limits, literally and figuratively, of various mathematical theorems. But they work for applied mathematicians, like actuaries, because they force theoretical mathematician authors to step out of their equations and notation for five-minute chunks of time and give concrete (counter) examples of what they mean by a theory they are explaining.[4]

But if you're a career actuary, you knew a long time ago you wanted to work in applied math, not pure math. Keep reading about statistical models, whether written by actuaries or economists or finance professors. Be author agnostic. And evangelize on what you learn. ■

Paul Conlin, FSA, MAAA, is an actuary at Aetna. He can be reached at *conlinp@aetna.com.*

ENDNOTES

1 Dr. Snow's original differences-in-differences table can be found in *Mastering Metrics* by Joshua D. Angrist and Jorn-Steffen Pischke (Princeton University Press, 2015).

2 The reason this works mathematically is the Law of Conservation of Angular Momentum, one of the five great conservation laws of mechanical physics. Besides lacking modern observational and computational tools, Kepler lived in the pre-calculus era, so he was unable to express his observations in integral form. For a concise explanation, see *Continuum Mechanics* by A.J.M. Spencer (Dover Publications, 2004). Mechanical physics has lots of parallels to econometrics in that the equation (and type of equation) is sought first, then the calculations begin.

3 Barth, Michael M., Evan M. Eastman, and David L. Eckles. 2019. It's About Time: An Examination of Loss Reserve Development Time Horizons. *North American Actuarial Journal* 23, no. 2:143–168.

4 The masterpiece of probability counterexamples is *Counterexamples in Probability* by Jordan M. Stoyanov (Dover Publications, 2013).