SOCIETY OF ACTUARIES Forecasting and Futurism Section

ISSUE 11 | JULY 2015

Forecasting & Futurism

NEWSLETTER

Why Smart People Make Big Money Mistakes

Review by Ben Wolzenski

Actuaries
Risk is Opportunity.®

ISSUE 11 | JULY 2015

Forecasting & Futurism

NEWSLETTER

2014-2015
SECTION LEADERSHIP
Chairperson
Doug Norris
FSA, MAAA, PhD
New Member Coordinator
doug.norris@milliman.com

Vice Chairperson
Brian Holland
FSA, MAAA
brian.holland@aig.com

Secretary/Treasurer
Ricky Trachtman
FSA, MAAA
ricardo.trachtman@milliman.com

COUNCIL MEMBERS
Geoffrey Hileman
FSA, MAAA
ghileman@kennellinc.com

David Snell
ASA, MAAA
dave@ActuariesAndTechnology.com

Qichun (Richard) Xu
FSA, PhD
Education Coordinator
rxu@rgare.com

Haofeng Yu
FSA, CERA, MAAA,PhD
hyu@aflac.com

Sheamus Kee Parkes
FSA, MAAA
sheamus.parkes@milliman.com

Bryon Robidoux
FSA, CERA, MAAA
brobidoux@rgare.com

BOARD PARTNER
Ian G. Duncan
FSA, FIA, FICA, MAAA
duncan@pstat.ucsb.edu

NEWSLETTER EDITOR
David Snell,
ASA, MAAA
dave@ActuariesAndTechnology.com

SOA STAFF
Sam Phillips, Staff Editor
sphillips@soa.org

Andrew J. Peterson, FSA, FCA, MAAA, EA, Staff Partner
apeterson@soa.org

Jennifer Foster, Section Specialist
jfoster@soa.org

Julissa Sweeney, Graphic Designer
jsweeney@soa.org

Published by the Forecasting and Futurism Section Council of the Society of Actuaries

This newsletter is free to section members. Current issues are available on the SOA website *(www.soa.org).*

To join the section, SOA members and non-members can locate a membership form on the Forecasting and Futurism Section Web page at *www.soa.org/forecasting-futurism.*

This publication is provided for informational and educational purposes only. The Society of Actuaries makes no endorsement, representation or guarantee with regard to any content, and disclaims any liability in connection with the use or misuse of any information provided herein. This publication should not be construed as professional or financial advice. Statements of fact and opinions expressed herein are those of the individual authors and are not necessarily those of the Society of Actuaries.

© 2014 Society of Actuaries. All rights reserved.

FROM THE EDITOR:

Forecasting & Futurism– "One of the Best Kept Secrets of the SOA"

By Dave Snell

A few days prior to starting this issue's introduction, I was in Chicago at a Society of Actuaries Cultivate Opportunities Team meeting, and heard an SOA Board member describe the Forecasting & Futurism section as "one of the best kept secrets of the SOA." It was meant mostly as a compliment; but also as a regret regarding our section name. The SOA has formed a major initiative to promote actuaries as the logical choice when any industry wants professionals to do predictive modeling (PM). Yet, most SOA members have no idea what our Forecasting & Futurism section does, and they are quite surprised to learn that we have been presenting sessions on, and publishing articles on, PM and related topics continually for at least the last six years. I'm even told that some SOA members assume we just sit around with tinfoil hats and talk about science fiction. Last year, the confusion increased a bit when the Modeling section was formed. Another member of the cultivate opportunities team, who heads up a PM department at a major insurer, said she joined the modeling section thinking it was where she would interact with the other predictive modelers. Then she discovered it was not for PM, but focused more on usage and controls for traditional actuarial models—and she had no idea the Forecasting & Futurism section had the PM focus she wanted!

Clearly, we have a perception issue. We are the section most interested in predictive modeling and predictive analytics, but also the section that recognizes the complexity of forecasting includes not just analytic models but also behavioral economics and other non-quantitative approaches to predictions that complement the strictly numbers approaches. It is difficult to convey our broad range of prediction techniques in a concise name, and thus, we sometimes suffer the tinfoil hat appellation due to faulty perceptions. Ironically, we are also the section that has published and presented the most on Behavioral Economics, which shows how perceptions can be so much more persuasive than facts. One of the sessions we scheduled for the Health meeting this year was *Predictive Analytics – The Reason Your Strictly Analytic Models Fail!*

Over and over we encounter situations where a mathematically sophisticated actuarial model will fail because it relies strictly on logic, and people just refuse to obey the rules of logic.

This issue includes some fascinating articles on behavioral economics. You can start with the chairperson's article from Doug Norris: "Can't Win for Losing." Doug tells us about the "Winner's Curse"—a phenomenon in which the winner of an auction, or the company with the lowest bid on a contract, may turn out to be more of a loser than a true winner. One interesting observation he makes is that "the greater the number of participants in the auction, the more likely it is that the ultimate winner has overvalued the item." Doug reminds us that knowledge is the best defense when we are developing rates, and he offers good advice on checks to make so that we do not suffer the winner's curse in our quest for a win.

Ben Wolzenski continues the behavioral economics lessons with his poignant review of *"Why Smart People Make Big Money Mistakes and How to Correct Them: Lessons from the New Science of Behavioral Economics."* Ben gives a

CONTINUED ON **PAGE 4**

"ONE OF THE BEST KEPT SECRETS" … | FROM PAGE 3

walkthrough of some of the chapters in this book by Gary Belsky and Thomas Gilovich, and summarizes some non-intuitive maxims such as "Not All Dollars Are Created Equal"—why casinos use house money so we will gamble more, "Herd it Through the Grapevine"—how investors follow the herd when it is too late, and "Dropping Anchor"—a stated number or name may have zero relevance to a situation; but it can bias our actions in that situation.

Tyson Mohr contributes another behavioral economics book review. He tells us about *Thinking, Fast & Slow,* by Daniel Kahneman. If you don't have the time to read the 500+ page book, read Tyson's review of it for an excellent summary in far fewer pages. The examples he draws from the book include such gems as "repeated statements become increasingly more believable and likeable" or as we might say "repeated statements become increasingly more believable and likeable." The mere exposure effect is difficult to understand; but it works. It really does. It often does work.

Does this mean there is no place for the quantitative techniques that actuaries know and love when it comes to judgement and forecasting? No! Fortunately, Mary Pat Campbell reassures us that they can be used to improve decisions over those made by purely qualitative methods and over those made simply by the so-called wisdom of crowds. In "What I've Learned from the Good Judgement Project," Mary Pat describes a new type of group prediction method. In this twist on crowd wisdom, the participants and their contributions are tracked, and the better predictors are accorded higher than average weightings for future group predictions. According to Mary Pat, this is a government funded project under a department with a surprising name: the Office of Anticipating Surprise. Read about the project, and Mary Pat's experience with it in her enlightening article.

Getting back to numbers, we also have a lot in this issue on various forms of predictive analytics. Brian Holland wrote an article on how we can deal with the real world problem of how to apply predictive models when your data is missing several values, or values are based on limited exposure. He writes about how to apply singular value decomposition (SVD) in these situations in his article "SVD of Weighted or Missing Data." In the first draft of Brian's article, he used the acronym WMD in the title, and my first impression was that these types of data gaps are weapons of mass destruction for the accuracy of our models. Brian arms us for battle with references to several academic tools and to an R programming package that helps us avoid the danger of overfitting our sometimes sparse data.

Shea Parkes and Brad Armstrong continue this line of discussion with an article that introduces us to a technique known as ridge regression. Ridge regression, as they show with an applied example, is especially useful when you have parameters and coefficients for a large population, with high credibility for that population; but you wish to adjust the coefficients that will be credibly different for a smaller, target population. Shea and Brad use a penalized regression and cross validation approach to choose a reasonable balance between standard weights from the larger population and completely retrained weights from the target population. Read their article, "Calibrating Risk Score Model with Partial Credibility" for the details and see how this approach can help you recalibrate your predictive analytics model for a moderate size, but not fully credible, target population.

Admittedly, Brian, Brad and Shea have contributed approaches for more experienced PM actuaries. What do we have in this issue for the actuary starting out with PM? Lots! Next in this issue we present an article with an unusual title: "Appendix B: How to Build a Model." This actually is a copy of an appendix of a research paper sponsored by the SOA Committee on Finance Research. You can read the entire research paper: "Lapse Modeling for the Post-Level Period—a Practical Application of Predictive Modeling" at the SOA site *https://www.soa.org/Research/Research-Projects/Finance-Investment/lapse-2015-modeling-post-level/#sthash.W9lERSls.dpbs,* but you can read this valuable appendix copied here in this issue to see how to build a PM step-by-step using your data and the R programming language. We thank Richard Xu, Dihui Lai, Minyu Cao, Scott Rushing, and Tim Rozar for their excellent paper and the Society of Actuaries, for the permission to reprint this portion of the paper.

Again, for the reader seeking a way to get started in the forecasting field, we have an article from Doug Norris, titled "Simple Rating Systems: Entry-Level Sports Forecasting." Doug is an avid sports fan. He would sometimes apologize for background noise on F&F council calls when he was still at the hockey rink. Although he sports a Ph.D. in mathematics, in this article Doug walks the reader from a very basic sports prediction algorithm: "when an undefeated team plays a winless team, the undefeated team usually wins" and gradually layers on levels of increasing sophistication without resorting to calculus, advanced statistics, or any Greek letters. This one, you can read without having to drag out study notes or your old textbooks. It's a winner!

Still, some of us like to make that leap to more advanced PM, but without the angst and wheel spinning often associated with self-study. Bryon Robidoux summarizes his experience at the Predictive Analytics World (PAW) conference this spring in San Francisco. Bryon's article, "Stepping Out," is one actuary's perspective on the value of a conference that might not be covered by your company; but still might be a prudent investment in your future if you wish to enter the PM field. Bryon describes a hands-on introductory class on using R for predictive modeling, summarizes keynote speeches from PM experts, and lists the PAW conference recommendations on how to prepare yourself for a PM position. One surprising item on the recommendation list was YouTube instructional videos. Another topic of interest was the one on the qualities of a good data scientist, which Bryon notes as very similar to those for a good actuary.

Speaking of Data Scientists, our non-actuary Data Scientist Friend of the F&F Council, Jeff Heaton, contributed his article "What Big Data is, and How to Deal with It." Jeff is a prolific writer for F&F and he is the author of several books on PM and related topics. His current series, *Artificial Intelligence for Humans* will see Volume 3 published later this year. Search on Amazon for "neural networks" and one or more of Jeff's books is likely to top the result list. In this article, Jeff describes the history and in some respects, the future of Big Data, and the tools we can use to handle it for PM and for machine learning. One such tool is Vowpal Wabbit, which sounds like something from a Bugs Bunny and Elmer Fudd cartoon ("Dwat that wabbit"), but it really is a popular approach to process a dataset of any size, as there is no need to load all the data into memory.

Big data is forcing us to enhance many of our tools. So is the increased actuarial usage of stochastic-in-stochastic analyses (nested stochastic processes), which can result in unacceptably long program run times. Many of our popular computer languages are not inherently well suited for parallel computations. This creates a bottleneck in an age where hardware costs have decreased dramatically and multiple machines may be cost effective but the software can't take advantage of them. Charles Tsai writes about a free and open source language solution from MIT named Julia. His article, "A 'Hot Date' with Julia: Parallel Computations of Stochastic Valuations," introduces us to Julia, and shows a four CPU example that runs significantly faster than the traditional non-parallel approach used by R and many other languages. In his discussion of whether Julia is a disruptive innovation, Charles gives an unbiased summary of both advantages and weaknesses of Julia for actuaries. His writing style takes a topic with the potential to be tedious and he makes it fast-paced and interesting. Whether you are ultimately interested in Julia or not, his discussion of the advantages of parallel processing is worthwhile to read.

I'm ending this issue with a summary article that is probably long overdue. We get a steadily increasing number of queries from actuaries asking how to get started with predictive modeling, behavioral economics, Delphi studies, genetic algorithms, machine learning, complexity sciences, classification and regression, etc. and over the past six years, the F&F

CONTINUED ON **PAGE 6**

"ONE OF THE BEST KEPT SECRETS" … | FROM PAGE 5

section newsletter has published more than 100 articles that touch upon these and other topics. Our last article for this issue is a list I've compiled of all these articles, the authors you might wish to contact for more information, and a very brief description of the article. I hope you find it helpful.

I mentioned the allusion to tinfoil hats already and repeating that phrase that is probably risky because, as we know from behavioral economics, "repeated statements become increasingly more believable"; but there is a part of F&F where we can proudly display the tinfoil hats: as part of our interest in Futurism we are cosponsors of the annual Actuarial Speculative Fiction contest. I have had the honor to be one of the judges for several years now and I personally look forward to each year's new collection of actuarially-related short stories of what the future may hold for us. You can read all 16 stories, including the overall winner (*Life After Death* by Ken Feng) and the F&F section winner (*Hotel Zukunft: The Future is Different* by Craig DeAlmeida) at the SOA website page *https://www.soa.org/Professional-Interests/2015-speculative-fiction-contest-final.aspx#sthash.h2Yhrlr3.dpbs* and I recommend them as thought provoking and enjoyable reads. In April of this year, Vanessa Drucker wrote an article about this contest in the *Retirement Income Journal* (Vol.301, April 23, 2015) titled, "Who Knew? Actuaries Have Two-Sided Brains."

As you can see from this introduction, the current issue is loaded with salient articles on predictive modeling, behavioral economics, big data, and new forecasting tools such as hot new programming languages. Read and share these ideas with your colleagues. We may currently be one of the best kept secrets in the SOA; but you have our permission and encouragement to share the secret. Hey, have you heard how cool the F&F section is? It's kind of a secret, but please pass it on!

Dave Snell

Dave Snell, ASA, MAAA, is technology evangelist at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *dave@ActuariesAndTechnology.com*

CHAIRSPERSON'S CORNER:

Can't Win for Losing

By Doug Norris

UCLA Bruins football coach Henry "Red" Sanders once said, "winning isn't everything; it's the only thing." Although in the sporting realm winning is generally considered a good thing, Sanders's advice does not necessarily translate into the real world (or even the actuarial world).

Picture a scenario where you work for a car dealership, and are attending an auto auction. You see a car that you like, and have a price in your head that you feel would be fair. The auction begins, and the bid price quickly rises above your expectations. Finally, the bidding slows, and you can win the auction with just one more bid. Alternatively, you must go back to the dealership empty-handed. You place your bid, at a higher value than you originally would have liked, and win the auction. But was winning the right move?

Back at the dealership, the car is placed on the lot for sale. Many months pass, and no one will buy the car at a profitable level. Winning the auction has proven to be very costly. We all have different estimates of what things are worth, and in an auction situation, the winning bidder will always have assumed the highest value. This phenomenon is known as the "winner's curse." The greater the number of participants in the auction, the more likely it is that the ultimate winner has overvalued the item. Moreover, the likelihood that the difference between the winning bid and the "fair" value will be large increases with the number of bidders.

Bidding situations exist throughout our world, with companies bidding for engineering contracts, water rights, or advertising space. In these situations, bidders typically assign different "values" to the object up for bid; this can be due to differing objectives, prior information, influences, stakeholders, and importance to the overall business. In the insurance industry, bidding is commonly done by setting rates to attract consumers and gain market share.

When the Patient Protection and Affordable Care Act (ACA) legislation was passed in 2010, there were a lot of things that we knew were unknown about the ACA market (and even some things that we didn't know that we didn't know). The relative morbidity of the then-uninsured commercial population, the level of pent-up demand that would be experienced in the post-ACA marketplace, the impact of future regulation and legislation, how well the exchanges would function, and the efficacy of the "3 Rs" risk mitigation programs all represented items that had an innate unknowability to them. Actuaries had various opinions on how they would impact prices. Ultimately, health insurers are "bidding" for customers, and in this case, the lowest price goes a long way toward determining who "wins" (attracts the most customers).

At the time this article is being written, we are in the midst of a third year of ACA pricing, and there are many things that we do not yet know. When bidding for customers in the commercial health care marketplace, it is important that carriers ensure that their assumptions focus on long-term sustainability and not just short-term market share gains. Yes, someone will have the lowest price in the market at the end of the day, but remember that market share is not the be-all and end-all. Even once rates are set and customers are gained, it's important to continually monitor emerging experience to test the validity of assumptions. Are claim costs coming in higher than I expected? Are more members focused in one product, or one area, or one demographic segment? Why? Tightening and honing assumptions benefits us all.

On the other hand, just because an insurer happens to offer the lowest price does not mean it is underpriced. First of all, someone has to offer the lowest price. One company may have true advantages in terms of administrative costs, better negotiated discounts with providers, or more efficient medical management practices. But the products of another might line up better with enrollees in terms of health outcomes (and also reduce their costs of care). That company may have estimated costs appropriately, whereas the others' estimates were too high.

This effect applies to actuaries in fields outside of health insurance as well. Consumers shop around for the cheapest auto, home, or life insurance policies, and there's always the

CONTINUED ON PAGE 8

CAN'T WIN FOR LOSING | FROM PAGE 7

risk that the winner's curse will rear its head. Remember to keep this phenomenon in mind when developing rates. As always, knowledge is the best defense. Sensitivity-test your assumptions. Know where breaking points are. Understand the markets. Get corroboration where possible. Make sure to be fully up to speed on any Actuarial Standards of Practice that may be relevant to setting assumptions. If risks are material, remember that it's OK to be conservative.

One place where I'll always feel like a winner relates to the three years that I've spent on the Forecasting and Futurism Section Council. My term as section chair will end at the SOA annual meeting in October, and it's been exceptionally rewarding. As a health actuary, I'm quite familiar with techniques that health actuaries use, but being exposed to thought leaders in a variety of practice areas has been a boon to my career. As the "tools and techniques" section within the SOA, this is a great place to stay abreast of innovative practices that can be applied broadly, and our section newsletter is a great resource. In this issue, Dave Snell has catalogued the history of the Forecasting & Futurism Section's newsletter articles, and I hope that you'll find it as useful as I do. There's just so much stuff out there, and this makes it a lot easier to find (and may also spur you on to write something in the near future).

Our volunteers, from within both the SOA and the industry, have been great at helping us spread the word about predictive modeling, behavioral economics, futurism techniques, analytics, complexity science, and other leading-edge topics for actuaries. If you're enjoying this issue, I'd invite you to consider writing an article, attending sessions at the SOA meetings, or volunteering within the section in some other way. Our section members are our lifeblood. From getting the chance to meet with many of you, I know that there are still a lot of untapped ideas out there. Where should we head next? It's up to you. If you'd like to talk more about areas where you might be able to best contribute, feel free to send me a note at the address below.

Congratulations to both Kurt Wrobel and Steve Mathys, who were chosen as co-winners of the 2014 Forecasting & Futurism iContest. As you can tell by the name of the contest, it took us a while to get to this point, but I think that each entry presents a compelling narrative (and ultimately, given the nature of the two finalists, it was tough to compare them against one another).

Also, congratulations to Craig DeAlmeida, the Forecasting & Futurism Section's winner of the 11th Annual Speculative Fiction contest. We were a co-sponsor (along with the Technology Section and the Actuary of the Future Section), and were very impressed with the slate of entries in this year's contest. I found it interesting (and truly telling) that there were six distinct prizes chosen, and that no single entry won more than one category. Be sure to check out the entries on the SOA's website, and consider entering next year's contest.

Last but not least, congratulations to the winner of the 2015 Forecasting & Futurism NCAA Bracket Challenge. Although we set up the contest at the (relative) last minute, we had 39 entries, but none fared better than SOA staff fellow David Schraub. David's bracket came on strong in the later rounds, and correctly predicted that Duke University would defeat the University of Wisconsin. In the interest of full disclosure, my bracket entry placed 36th out of 39 entries. Please do not correlate this showing with my article on successful sports forecasting, later in this issue.

Anyhow, please enjoy the newsletter—with each issue, Dave Snell sets the bar even higher, and I think that we're at Renaud Lavillenie levels now (current world record pole vaulter). Hopefully, you can help us to set the bar even higher next time!

Doug Norris

Doug Norris, FSA, MAAA, Ph.D., is a principal and consulting actuary at Milliman in Denver, Colo. He can be reached at doug.norris@milliman.com.

Why Smart People Make Big Money Mistakes

Why Smart People Make Big Money Mistakes and How to Correct Them: Lessons from the New Science of Behavioral Economics *by Gary Belsky and Thomas Gilovich, reviewed by Ben Wolzenski*

The authors explain: "traditional [economic] theory holds [that] we make decisions because of a consistent and sensible pursuit of satisfaction and personal fulfillment, of getting the most out of life with our current and future resources." They then proceed to give examples of how this is not the case in reality, and posit behavioral economics as a way to explain why not. Early research (1970s) by Tversky and Kahneman showed that people use heuristics rather than logically thinking through every decision. Human heuristics developed over millennia are generally useful, but not always; "… in some ways, behavioral economics can be fairly described as the study of obsolete heuristics."

CHAPTER 1—NOT ALL DOLLARS ARE CREATED EQUAL

The authors start with an anecdote that illustrates why we do not value "house money" (at a casino) equal to other money. This is the "inclination to value and handle money differently depending on where it comes from, where it is kept, or how it is spent." This is why "Reimbursements send people on trips to the bank. Bonuses send people on trips to the Bahamas."

"Mental accounting" is one of the pillars of behavioral economics. It explains why it is worth extra time and effort to buy an item at 50 percent off for $25 instead of $50, but not worthwhile to exert the same extra time and effort to save the same $25 off the price of a $500 item, only a 5 percent savings. Or why a small bonus or refund is more likely to be spent ("found money") than a large bonus or refund, which is more likely to be saved. Credit cards also cause us to treat dollars differently. In a "landmark" experiment, bids for prime tickets from the half of bidders who were told the high bidder would pay by credit card averaged about twice as much as those from the half who were told the payment would be in cash. But mental accounting can be used to one's advantage, too. Certain workers who were paid weekly found it hard to save, but when they were paid in six envelopes, one for each day they worked and a sixth not tied to any day, their savings increased fourfold within three months.

CHAPTER 2—WHEN SIX OF ONE ISN'T HALF A DOZEN OF THE OTHER

Prospect theory is the second pillar of behavioral science. The name comes from an oft-cited 1979 article by Tversky and Kahneman: "Prospect Theory: An Analysis of Decision Under Risk." This chapter deals with two aspects—loss aversion and the sunk cost fallacy.

An example cited is described below.

- You are given a sum $X and the choice of
    1. accepting a sure additional $.5X (to end with $1.5X) or
    2. flipping a coin to determine whether you get nothing more or an additional $X, ending with either $X or $2X.
- Research says you are more likely to choose option 1, the sure gain.
- You are given a sum $2X and the choice of
    1. accepting a sure loss of $.5X (to end with $1.5X) or
    2. flipping a coin to determine whether you lose nothing or lose $X, ending with either $X or $2X.
- Research says you are more likely to choose option 2, the chance to lose nothing.

(It would be interesting to know if the research results would be the same if the subjects were all actuaries.) Since the outcomes are equivalent, traditional economics suggests that you would be no more likely to choose option 1 in the first case than the second. The authors describe the reasons for the different choices in these terms.

"Prospect theory offers an alternative approach. It says that people generally do not assign values to options based on the options' expected effect on their overall level of wealth.

CONTINUED ON **PAGE 10**

… Prospect theory says we assign values to gains or losses themselves. …It is the actual gaining or losing—and our feelings about it—that matters more to us. …"

This helps explain why individual investors typically sell their winning investments too soon (to lock-in a gain) but hold onto their losing investments too long (to avoid booking a loss). The authors cite research data that shows that is indeed what happens: the stocks that investors sold outperform those they held over the next year.

"Framing" refers to the specific environment and/or language in which questions or problems are presented. As an example, votes regarding a proposed school tax increase were more favorable when the polling location was situated … in a school.

The authors relay several anecdotes to exemplify the "sunk cost fallacy," including the results of this experiment. Discounts were randomly distributed to theater subscribers. The result was that subscribers who paid more for their tickets (greater sunk cost) attended performances more often than those who received discounts, even though everyone had initially expected to pay full price.

Applying these lessons to investments and financial management generally, the authors enumerate 12 "suggestions that should help you make wiser decisions." These are written in a reader-friendly (not technical) narrative—and I think they represent good advice.

CHAPTER 3—THE DEVIL THAT YOU KNOW

What causes "decision paralysis?" Research shows that decisions to delay or take no action are more likely "when there are many attractive options from which to choose." Needless to say, not taking any of a number of attractive financial options will usually produce worse results than choosing one. Research further showed that the greater the number of attractive choices, the more likely it was for no choice to be made. The extent of choice difficulty depends on the extent to which a person is a "maximizer" (one who wants the best) rather than a "satisfier" (one who wants "good enough").

The book has a few inset boxes with their own narrative. There is one in this chapter; the following is a direct excerpt. "BIG EYES—Options are con artists. They seduce with a promise of joy but often leave us confused and wanting. Consider this experiment … When they offered consumers a choice of different digital devices … some six in ten picked the option with the most features. … But when actually using their new gizmos, most consumers quickly fell prey to … "feature fatigue"; that is, they quickly tired of using all those extras (if they even figured out how to). … We might just say that humans have "big eyes. …"

The huge number of investment choices (over 8,000 mutual funds, plus individual securities and ETF's) encourages investment decision paralysis. In employer sponsored plans, a research study showed that employee participation rates decreased 2 percent for every 10 additional investment choices added.

A related phenomenon is "status quo bias." In an experiment, a group of students with finance experience were given a choice of four investments with different degrees of risk and potential return. With a clean slate, the distribution of choices was 18 percent-32 percent-32 percent-18 percent from most to least risky—a nice bell curve. However, a different result was obtained when the question was presented as a large amount is already invested in one of the same four choices and how would you choose to deploy it. No matter which of the four existing investments was the current place for the investment, that investment was the most popular choice for future investment!

In "What's Mine is Mine, and What's Yours isn't Worth as Much" describes the "endowment effect." If a person owns something, the sale price is significantly higher than the same person would be willing to pay for the same something. Then there's "regret aversion"—we'd rather feel bad for something we didn't do than for something we did, even if the net result is the same.

The authors once again conclude the chapter with a set of suggestions—seven this time.

CHAPTER 4—NUMBER NUMBNESS

One can sum up this chapter's theme by the authors' observation that "… people have trouble with numbers." This may not be true of actuaries and a few other professions of our ilk, but the book describes how this is too often true of the general population in an entertaining manner.

The authors identify and exemplify three of the ways this leads to money mistakes: not taking inflation into account; mistaking or misusing probability ("Odds Are You Don't Know What the Odds Are"); and a bias toward bigness (people tend to discount the importance of small numbers, such as small but frequent expenses). Even though some of the financial examples reflected the book's 1999 vintage, this was an easy and enjoyable chapter to read.

Shown below are this chapter's suggestions to avoid "big money mistakes" with some explanatory additions win brackets.

- "Don't be impressed by short term success [of investments]."
- "Because chance plays a far greater role than you think in investment performance, you should play the averages."
- "Know when time [and compound interest] is on your side and when it isn't."
- "Enhance the base rate [mind long term trends]."
- "Read the fine print."

CHAPTER 5—DROPPING ANCHOR

In this chapter the authors explain and provide ample evidence for "anchoring" and "confirmation bias." Anchoring is defined as "clinging to a fact or figure or idea that may or may not have relevance to your judgments or decisions." Confirmation bias is "a tendency to search for, treat kindly, and be overly impressed by information that confirms your initial impressions or preferences." As usual, the authors' examples bring these terms to life, as does their turn of phrases, such as: "Once an idea sets in your head, it often sets in concrete; you can break it, but you may need a sledgehammer." Anchoring can work for marketers in many ways, including the suggestion of how much to buy. (I once overheard a grocery store employee tell another that they sold more of an item when it was advertised as "10 for $10" than when it was advertised with a $1 price!) Sure enough, the authors cite similar examples. Anchoring can result from numbers that have nothing whatsoever to do with the value in question. "In another study, participants were asked what they were willing to pay for a meal at restaurant 'Studio 97' or 'Studio 17.' We're sure you can guess the result. On average, participants were willing to pay one-third more for a meal at 'Studio 97'." The chapter concludes with five suggestions for avoiding bad decisions due to anchoring and confirmation bias.

CHAPTER 6—THE EGO TRAP

The general meaning of this chapter can be inferred from the title, or from this line in the chapter: "… almost as long as psychologists have been exploring human nature, they have been amassing evidence that people tend to overestimate their own abilities, knowledge and skills." What is impressive is the evidence provided. Study after study and example after example demonstrate that this is true and is broadly based. That is, it applies to people of widely different backgrounds and levels of knowledge or skills. One of my favorite studies is about how consistently people underestimate how long a task or project will take.

When it comes to financial decisions, the implication is that people think they are in better financial condition than they are. So people are often underprepared for what lies ahead, and too often willing to make substantial spending decisions while not as well informed as they think they are. What does this mean about investment decisions? The authors contend that most people "have no business at all trying to pick investments, except perhaps as sport" and cite supporting research.

CONTINUED ON **PAGE 12**

WHY SMART PEOPLE … | FROM PAGE 11

Investor overconfidence is maintained in the face of not-always-supporting experience by a phenomenon described as "heads I win, tails it's chance." We tend to attribute success to our ability, failure to bad luck. This chapter's concluding advice includes paragraphs on "Investor, Know Thyself," "Ask Three Good Questions" and "Get a Second Opinion."

CHAPTER 7—HERD IT THROUGH THE GRAPEVINE

The theme here is that retail investors "follow the herd" when it's too late. They buy investments after they have done well and sell investments after they have done poorly. The authors cite a single factoid that shows this all too clearly: from 1988-2008 all stock and bond mutual funds averaged returns of 8.4 percent and 7.4 percent respectively, but the investors in stock and bond funds averaged 1.9 percent and less than 1 percent respectively! Other examples are cited to the same point.

The authors point out that there are a number of ways in which society encourages conformity—following the herd, in this case. The tendency to conform is enhanced in uncertain situations—such as choosing investments when one is not an expert in the area. Doing what everyone else is doing seems a reasonable choice.

The chapter-ending advice advocates patience, avoiding hot investments, establishing investment rules and sticking to them.

CHAPTER 8 – EMOTIONAL BAGGAGE

"… emotions are partners in all the decision-making processes we've been discussing. …" Different parts of the human brain produce emotions (the reflexive system) and logic (the reflective system). The authors assert that behavioral economics needs to take both into account, since emotion affects all behavior, including financial behavior. They support the assertion with page upon page of examples and research results. We feel better when the weather is nicer, but did you know that "examination of stock markets in 26 countries over a 15-year period revealed that the amount of sunshine on a given day is … positively correlated with market performance"? On the other hand, "When we feel bad … about one thing, significant or not, it can color our view of all things at that moment."

At chapter's end, the authors offer techniques for keeping emotions from having too much influence on our financial decisions, under these headings:

- Voice your reason;
- Use checklists;
- Play decision chess;
- Mind your pros and cons; and
- Don't just do something, stand there.

CONCLUSION & POSTSCRIPT

As a prelude to their 14 "Principles to Ponder" and eight "Steps to Take" the authors make this acknowledgement. "It is also difficult to alter many of the behavioral-economic habits we've discussed in this book because, although they cost you money, they reflect psychological tendencies that bring great benefits in other ways or in other areas."

The "Principles" and "Steps" provide an excellent summary of the conclusions and advice of the book, without all the supporting evidence. If you cannot spare the time to read the whole book, go to the Conclusion first; it is only 18 pages long. But when you can, go back to the full text to be entertained and further informed.

Ben Wolzenski

Ben Wolzenski, FSA, MAAA, is managing member at Actuarial Innovations, LLC in St. Louis, Mo. He can be reached at *bwolzenski@rgare.com.*

Thinking, Fast and Slow

Review by Tyson Mohr

As a reader of this newsletter, you've almost certainly heard of Daniel Kahneman's *Thinking, Fast and Slow.* You might have even picked up a copy for yourself, only to realize that this 500-page tome is not as much a book as a lifestyle choice. The depth and breadth of the content make a comparative summary challenging. Instead, I will share the experiments and conclusions that I found particularly compelling. (I will also include technical terms in parenthesis to facilitate further research if the topic interests you.) These examples will give you a sense of the content of the book, and hopefully encourage you to investigate further. The summary is divided into five parts, which align with the parts of the book.

Two housekeeping items:

I wrote this summary in second person in response to research (explained in the book) showing that information about people's behavior in general does not typically change your perception of how individual people will act. (See the "Helping Experiment" for an example.) The information is therefore framed as surprising facts about the reader, not surprising facts about all of us as humans. I certainly consider myself just as susceptible to these results as anyone else.

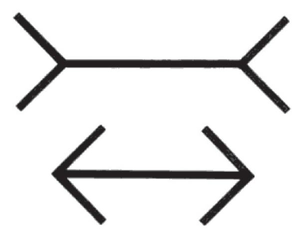
Also, I will often ask you to consider your intuitions about certain topics. Actuaries are trained to identify and apply techniques to avoid some of the pitfalls described. You will have more fun if you take off your actuary hat and consider how you engage in everyday life.

Part 1: Two Systems

The title of the book refers to the theory that mental activities can be roughly separated into two "Systems." System 1 thinks fast, with little or no effort, and sometimes without awareness. System 2 thinks slow, allocating attention to effortful mental activities. System 2 is also lazy—it only wants to think as little as possible to solve a problem. System 1 is constantly forming beliefs and conclusions, and System 2 does not usually challenge System 1 without deliberate effort. Here are some examples of the activities of each system.

System 1	System 2
Make a disgusted face when shown a horrible picture	Describe how to make a disgusted face
Recognize that someone has an angry expression	Monitor the appropriateness of your behavior in social situations
Answer 2x2=?	Answer 17x24=?

In the Muller-Lyer Illusion, shown below, you cannot help but perceive the top line as longer, even after you measure it. Cognitive illusions, mental processes which are predictably biased, persist in the same way. We are wise to learn to identify visual and cognitive illusions so we can employ methods to overcome them. (Cognitive Illusions)



* * *

If you like someone's politics, you probably also like their appearance as well. Evidence of positive characteristics make you inclined to view all characteristics favorably, and vice versa. (Halo Effect)

* * *

"Michelle is intelligent and strong. Will she be a good leader?" You probably think she is, since the stated qualities are desirable. But what if she is also manipulative and cruel? We make judgments based on available evidence without feeling a need to seek out more detail. (What You See Is All There Is)

CONTINUED ON PAGE 14

* * *

Repeated statements become increasingly more believable and likeable. Repeated statements become increasingly more believable and likeable. (Mere Exposure Effect)

* * *

The answer to "How happy are you with your life?" is strongly influenced by how happy you are at that particular moment. When asked a hard question, you sometimes answer an easier question without even knowing it, like "How happy am I now?" (Substitution)

Part 2: Heuristics and Biases

Kahneman was particularly interested in the degree to which people do or do not (mostly do not) have an intuitive understanding of probability and statistics. He identified numerous consistent flaws in reasoning (Biases) and problem-solving approaches (Heuristics).

The lowest concentrations of cancer are in rural states. You can easily construct a causal story to explain this (e.g., cleaner environment). However, the highest concentrations of cancer are also in rural states. You can again come up with a causal story (e.g., poor access to health care). Both cannot be true. The real story is that there are fewer people in rural states, and outliers are more likely when a sample set is small. (Law of Small Numbers)

* * *

Two groups of people were asked how old Gandhi was when he died. But before they were asked to guess, Group A was first asked if he was over 114, while Group B was asked if he was under 35. Group A guessed considerably higher than Group B. Even though the initial numbers were transparently unreasonable, they nevertheless influenced the guesses. If you think those marked down prices on the sale rack don't impact your purchasing decisions, think again. (Anchors)

* * *

Earthquake insurance purchases increase after earthquakes. This is because when asked about probability, you actually assess how easy it is to think of specific instances. The problem is that things other than frequency determine how easy it is to recall instances, such as media coverage or personal experience. (Availability Heuristic)

* * *

Google the "Linda Problem." You will probably share the intuition that Linda is more likely to be a feminist bank teller than just a bank teller. This is impossible since the former is a subset of the latter. Your intuition is due to the fact that more detailed descriptions often seem intuitively more likely. (Less is More)

Part 3: Overconfidence

The test of an explanation is whether it makes events predictable in advance, not whether it can explain past events. Prediction error is inevitable because the world is unpredictable. Short-term trends can be forecasted and behaviors can be predicted from the past, but a success rate of 20 percent is excellent. There is no harm in attempting forecasts, but it is dangerous to be overly confident in them.

There is great demand for two genres of business writing: the history of the rise/fall of a company, and an analysis of differences in successful and unsuccessful companies. After the books are written, most of the chronicled successful companies perform worse. These stories typically ignore the role of luck, which is involved in most great successes. It's difficult to develop skill at creating successful companies because one can only make a small number of attempts in one's lifetime, and the reason for failure or success is often not apparent. (Narrative Fallacy)

* * *

When you buy a stock, who sells it, and why? Both sides have an illusion that they have better information, even though the evidence clearly shows that the performance of the most active traders is no better than random. Yet investors feel as if they are exercising skill, and when evidence conflicts with personal experience, the evidence is ignored. Remember: unless there's specific evidence that you're not average, you most likely are. (Illusory Skill)

* * *

Seasoned surgeons have extremely accurate, almost magical, intuitions about when a patient is at risk during a surgery. However, their ability to forecast recovery time is no better than random guessing. What is the difference? The fact that surgery has the following characteristics which allow for the development of skilled intuition, whereas long-term forecasting does not:

- A somewhat predictable environment
- Regular feedback on success/failure
- The opportunity to learn regularities through prolonged practice (about 10,000 hours, or five years of 40 hours/week)

(Low-Validity Environment)

* * *

Kahneman planned to complete a project in 2 years. His plans were unaffected by information that in similar projects only 40 percent succeeded and those who did took 7 to 10 years. His project ultimately failed after 8 years. When you plan, you prefer to see yourself as a special case (Inside View) instead of part of a reference class (Outside View). You also tend to prefer best-case to realistic assessments (Optimism Bias) and to keep investing in lost causes to avoid admitting failure (Sunk Cost Fallacy).

Part 4: Choices

This part deals with Kahneman's contributions to the field of Decision Science. This research made him one of the founders of Behavioral Economics and won him the Nobel Prize in Economics.

Today, Jack and Jill each have wealth of $5 million. Yesterday, Jack had $1 million and Jill had $9 million. Are they equally happy? Of course not—Jack is ecstatic and Jill is distraught. The absolute value of wealth does not determine your happiness as much as your change in wealth from your previous reference point. (Reference Points)

* * *

Consider each of the choices below independently.

1. You are given $1000. Choose to get $500 for sure or 50 percent chance to get $1,000.

2. You are given $2000. Choose to lose $500 for sure or 50 percent chance to lose $1,000.

You probably prefer the sure thing in Choice 1, but prefer the gamble in Choice 2. You tend to be risk averse when facing two gains and risk accepting when facing two losses. Yet the choices are mathematically identical. The sizes of initial gifts aren't appropriately taken into account by our intuitions. They become reference points, and as such are treated as valueless. (Prospect Theory)

* * *

Participants were given one of two gifts of approximately equal value (a pen or a chocolate bar) and asked to fill out a questionnaire. After they completed the questionnaire, they were asked if they wanted to trade their gift for the opposite one. Only about 10 percent switched. Ownership has intrinsic value. (Endowment Effect)

* * *

"Mr. Brown almost never picks up hitchhikers. Mr. Smith frequently does. Today both of them picked up hitchhikers and were robbed. Who will experience greater regret? Who will be criticized more severely?" You are likely inclined to believe that Mr. Brown will experience more regret, yet Mr. Smith is more blameworthy. Regret comes from taking a specific action out of character. Social blame comes from acting out of the norm. (Hindsight Bias)

* * *

For a certain health insurance plan, smokers pay 20 percent more than non-smokers. Should this be explained as a surcharge on the smokers or a discount for non-smokers? The psychological difference between these two framings is significant. Richard Thaler and Cass Sunstein's book *Nudge* dives into the practical applications of these types of considerations, as well as many other parts of Kahneman's work. (Framing)

CONTINUED ON **PAGE 16**

THINKING FAST AND SLOW | FROM PAGE 15

Part 5: Two Selves

- Would you go on a vacation if you knew your memory and all evidence of it would be erased?
- Would you prefer a drug that dulled the pain 50 percent throughout a painful surgery, or one that merely caused you to forget that the pain occurred?

Separate from the two Systems are two Selves: the Experiencing Self and the Remembering Self. The former has sensations in the current moment, while the latter accesses and reflects upon those experiences. Kahneman studied situations in which the interests of these Selves were put at odds with each other. In most cases the Remembering Self's interests trump the Experiencing Self's. You strive to maximize memories of experiences, not actual experiences.

You must place your hand in uncomfortably cold water for a period of time. Which of the following do you prefer?


Tyson Mohr

A. 60 seconds at 14° (Celsius)
B. 60 seconds at 14° and 30 seconds at 15°

Although B is clearly worse, you should choose B, since the slightly warmer water at the end will lead you to have a more favorable memory of B. (Peak End Rule, Duration Neglect)

Tyson Mohr, FSA, is an ERM project manager over model validation at State Farm. He can be reached at tyson.mohr.prm7@statefarm.com.

* * *

Does money buy happiness? Research shows that being poor makes you miserable, but being rich does not on average improve well-being. The most influential factors for happiness include how often you spend doing activities you would rather continue (Flow) and how much time you spend with people you love. (Experienced Well-Being)

* * *

What proportion of a day do paraplegics spend in a bad mood? You're inclined to say a fairly large percentage, but there is actually no difference from the general population. Paraplegics become less happy when they focus on their condition, but for most of their life they adapt and have positive and negative experiences just like anyone else. (Focusing Illusion)

CONCLUSION

A summary can hardly do the book justice, but I hope this has inspired you to read more. *Thinking, Fast and Slow* is influential, entertaining, and potentially life-changing.

What I've Learned from the Good Judgment Project

By Mary Pat Campbell

We have often heard of the supposed wisdom of crowds, and the downfall of experts, but as one person noted last year, not all crowds are all that good at predicting:[1]

> *"I read the results of my impromptu experiment as a reminder that crowds are often smart, but they aren't magic. Retellings of Galton's experiment sometimes make it seem like even pools of poorly informed guessers will automatically produce an accurate estimate, but, apparently, that's not true."*

The context of that quote was that the author, Jay Ulfelder, had a cousin who ran an impromptu contest online, asking people how many movies he had watched (in the theater) in the past 13 years. The cousin kept a record of every movie he watched (to remind himself of the perk of being master of his own schedule as a freelance writer).

Forty-five people submitted answers, and the average (the supposed "wisdom of crowds") was way off from the actual answer. However, some of the answerers were close to the true answer.

Jay continues:

> Whatever the reason for this particular failure, though, the results of my experiment also got me thinking again about ways we might improve on the unweighted average as a method of gleaning intelligence from crowds. Unweighted averages are a reasonable strategy when we don't have reliable information about variation in the quality of the individual guesses (see here), but that's not always the case. For example, if Steve's wife or kids had posted answers in this contest, it probably would have been wise to give their guesses more weight on the assumption that they knew better than acquaintances or distant relatives like me.
>
> Figuring out smarter ways to aggregate forecasts is also an area of active experimentation for the Good Judgment Project (GJP), and the results so far are encouraging. The project's core strategy involves discovering who the most accurate forecasters are and leaning more heavily on them. I couldn't do this in Steve's single-shot contest, but GJP gets to see forecasters' track records on large numbers of questions and has been using them to great effect. In the recently-ended Season 3, GJP's "super forecasters" were grouped into teams and encouraged to collaborate, and this approach has proved very effective. In a paper published this spring, GJP has also shown that they can do well with nonlinear aggregations derived from a simple statistical model that adjusts for systematic bias in forecasters' judgments. Team GJP's bias-correction model beats not only the unweighted average but also a number of widely-used and more complex nonlinear algorithms.

What is this "Good Judgment Project" and who are their forecasters?

Jay's post happens to have been written at the end of their third season, and I've joined the GJP for the 4th season. While there are details of the current season I can't share, I can explain the background of the project, some of the basics of participation, and, most importantly, what I've learned so far.

HISTORY OF THE GOOD JUDGMENT PROJECT

The Good Judgment Project sprouted out of a number of surprises in the U.S. intelligence community. How could they have been blindsided by so many developments?

Part of the research coming out of those failures was a competition called the IARPA ACE tournament. IARPA stands for Intelligence Advanced Research Projects Activity, providing funding and running projects that are intended to dig into intelligence issues that cross multiple organizations within the U.S. government. According to their own description, IARPA undertakes "high-risk/high-payoff research … [in which] failures are inevitable. Failure is acceptable so long as the failure isn't due to a lack of technical or programmatic integrity and the results are fully documented."[2]

CONTINUED ON **PAGE 18**

WHAT I'VE LEARNED … | FROM PAGE 17

The Good Judgment Project feeds into that mission—especially for the individual participant. Failure is a big part of the project—failure in forecasting. But more on that in a bit.

The ACE tournament run by IARPA stands for "Aggregative Contingent Estimation," and it's run under the Office of Anticipating Surprise (man, I'd love to direct that office). It was an attempt to provide better forecasts of geopolitical events. The Good Judgment Project is a spinoff from the project, being run by researchers at University of Pennsylvania and UC-Berkeley. They had put together an approach, in which forecasters were trained and measured that outperformed many of the other ACE tournament participants, and IARPA ACE sponsored them as a part of a four-year research project.

What is interesting is that while the project has discovered "superforecasters" as part of their project, they have also shown effective ways to train people to forecast.[3] The training involves learning how to think probabilistically (which we actuaries should be good at), how to battle cognitive bias (which we may be no better than most people), and in general, how to become more successful in forecasting.

As noted in the article referenced above, there are "clusters" of questions that are attacking higher-level issues from different angles:

> "Within each cluster, we offer numerous specific forecasting questions. For example, within the cluster about European economic and political integration, we asked a question in fall 2014 about whether voters in Scotland would pass the independence referendum, and within the Iran cluster, we have a question currently open that asks when Iran will release Jason Rezaian, the Washington Post's Tehran bureau chief, who has been detained for over five months."

I have seen both questions, one obviously closed (the Scots did not vote for independence) and the other still open. I will not comment on the questions specifically, but about what I've learned about myself and about forecasting in general.

JOINING SEASON 4

I first heard about the Good Judgment Project via an NPR story in April 2014:[4]

> "For the past three years, Elaine Rich and 3,000 other average people have been quietly making probability estimates about everything from Venezuelan gas subsidies to North Korean politics as part of the Good Judgment Project, an experiment put together by three well-known psychologists and some people inside the intelligence community.
>
> "According to one report, the predictions made by the Good Judgment Project are often better even than intelligence analysts with access to classified information, and many of the people involved in the project have been astonished by its success at making accurate predictions.
>
> "When Rich, who is in her 60s, first heard about the experiment, she didn't think she would be especially good at predicting world events. She didn't know a lot about international affairs, and she hadn't taken much math in school.
>
> "But she signed up, got a little training in how to estimate probabilities from the people running the program, and then was given access to a website that listed dozens of carefully worded questions on events of interest to the intelligence community, along with a place for her to enter her numerical estimate of their likelihood. …
>
> "She's in the top 1 percent of the 3,000 forecasters now involved in the experiment, which means she has been classified as a superforecaster, someone who is extremely accurate when predicting stuff like: Will there be a significant attack on Israeli territory before May 10, 2014?
>
> "In fact, Tetlock and his team have even engineered ways to significantly improve the wisdom of the crowd—all of which greatly surprised Jason Matheny,

one of the people in the intelligence community who got the experiment started.

"'They've shown that you can significantly improve the accuracy of geopolitical forecasts, compared to methods that had been the state of the art before this project started,' he said.

"What's so challenging about all of this is the idea that you can get very accurate predictions about geopolitical events without access to secret information. In addition, access to classified information doesn't automatically and necessarily give you an edge over a smart group of average citizens doing Google searches from their kitchen tables."

At the end of the article, I noticed they were going to start recruiting people for the fourth round in the research. All the prior forecasters who wanted to continue would do so, but there would be a new crop of people coming in. I pre-registered and then qualified by taking an online quiz touching on a variety of geopolitical subjects (most news junkies can easily answer these) as well as some reasoning items.

After being accepted in the fourth season, I got some training, involving some big themes in putting together a forecast and in improving one's performance. I always have access to these materials if I need to review the concepts, but I knew several of these just due to my own readings on cognitive biases.

One of the most important things I learned, though, was how I'd be scored.

HOW TO EVALUATE A FORECAST … AFTER THE FACT

One of the most important parts of the project is that forecasts get a score for their forecast accuracy after the fact. What's used is a Brier score, originally developed by Glenn W. Brier in 1950. The GJP uses the full Brier score, originally developed by Glenn W. Brier in 1950,[5] which works for a wider variety of questions than the yes/no example given below. However, for the purposes of illustration, I'm going to use the simplest formulation of this score.

The simplest type of forecasting question is forecasting the probability that a specific event will happen. The outcome will be yes/no, and you're putting a probability on the "yes" occurring.

To give a really simple question: "Will it snow >1 inch in North Salem, N.Y. on March 1, 2015?"

Let's pretend I forecast this for the five days preceding March 1:

Date	Forecast Probability
24 Feb 2015	50%
25 Feb 2015	50%
26 Feb 2015	60%
27 Feb 2015	75%
28 Feb 2015	95%

At the end I will calculate a Brier score, which depends on whether I came close to the actual result, 0 = it didn't happen and 1 = it did.

It just so happened we got over an inch of snow on March 1.

If I had prescience, I would have predicted 100 percent for each day. That will be one extreme.

If I had the opposite of prescience, I would have predicted 0 percent for each day. That will be the other extreme.

The basic Brier score formula is:

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_N)^2$$

Where N is the number of days in the forecast period, f_t is the forecast percentage on day t, and o_N is the ultimate outcome.

This score was originally developed for weather forecasts, where one would make a prediction of probability of rain for each day—each day would have one forecast. The GJP is

CONTINUED ON **PAGE 20**

looking at something different—because this is about events possibly developing over time, one would want to see forecasts coalescing and changing over time. One would hope it gets closer to correct.

If I had perfect prescience, the Brier score would be 0, and if the perfect opposite, the result is 1. So the lower the Brier score, the better.

Date	Forecast Probability	Brier score for day
24 Feb 2015	50%	0.25
25 Feb 2015	50%	0.25
26 Feb 2015	60%	0.16
27 Feb 2015	75%	0.0625
28 Feb 2015	95%	0.0025
	OVERALL BRIER SCORE	0.145

In my example, I did not do too poorly. The Brier score for going with a 50/50 guess is 0.25, so one would compare against that. The Brier score used by the GJP is not exactly as I did above, because it needed to be adaptable to multiple-choice answers, and not merely yes/no. The specific details are not important.

One important thing to note: because of the squaring of the difference of the probability and the actual outcome, one is penalized for being far from the mark. If you way underpredict the chances of an event, you get heavily penalized; and if you overpredict the chances, you get extremely penalized. Deviating in the wrong direction from the 50/50 mark hits you very hard, so one must be spare with predictions of 5 percent or 95 percent probabilities for any event.

But the point is that once a question is closed, and all the GJP questions are of a finite period and do get resolved (more on that in a bit), one can look at how one did. More importantly, one can look at one's own rank among forecasters within a small group.

THE IMPORTANCE OF FEEDBACK IN FORECASTING

I have forecasted a few dozen questions that have closed thus far, having made more than 100 individual forecasts (one can change the forecast for any specific open question once a day). I am grouped with several other people, but that's only for comparison purposes right now. Unlike some of the prior seasons, we new forecasters are not being grouped to collaborate on forecasts. Yet.

In addition, unlike prior seasons in which forecasters were not expected to explain their reasoning behind their forecasts, we are encouraged (and given extra points) for flagging key comments about our reasoning, and also checking off categories of activities we did to make the forecast (such as adjusting a forecast for the passage of time—getting closer to a deadline may make the event more or less likely to be fulfilled.)

Finally, I can look at the closed questions and see how they were resolved, and look at my entire forecasting history for the question. I have found this the most valuable portion of the project for me. We are encouraged to write post-mortem comments for ourselves (which we can also look at later), and in doing these post-mortems I have discovered the following:

REGULAR UPDATING IS GOOD

Part of the reason my scores were bad on some questions was because I did not revisit my forecasts often enough. I do not have time to check every day, but I do make sure I look at all my forecasts at least once a week.

TRY MORE!

Originally, I stuck to areas where I understood the issues better (or so I thought), and I'm coming to realize that I'm losing some valuable points thereby. Most of us aren't experts in all the topics being covered, and just doing a few Internet searches can do enough to get one off the 50/50 line for a forecast.

This one I'm still having trouble with.

REMEMBER THAT NOT ALL TIME SERIES (OR QUESTIONS) ARE EQUAL

One question I messed up on was because I forgot how volatile certain time series can be. Some of the questions asked are based on financial markets indicators, and not all of them develop smoothly.

In particular, I have to be careful of "threshold" questions—some of the finance questions are whether a particular financial indicator goes above or falls below a particular threshold within a time period. That's a very different question from whether it will still be above that threshold at the end of the period.

I had forgotten that certain things could jump drastically on news, within a few hours even, and though the particular item I was following did settle back to "normal" areas, the threshold had been crossed. And my Brier score was hit.

STAY AWAY FROM FUZZY QUESTIONS

Most of the questions they've been presenting to forecasters are very clear: will a certain event occur by a certain date? Whether that event occurred is usually clear to all.

However, they've started getting into "fuzzier" questions, and I have found some of what is being done with those frustrating.

I understand why they're doing this—the really important intelligence would tend to be of a fuzzy nature. They are also trying to be fair—there are responses and clarifications. One can request a clarification while a question is still open. After questions close, you can provide feedback as to whether you agree with how they resolved a question.

That said, I have limited time. I do not need extra frustration in my life and enough fuzzy questions in my day job. Maybe, if the GJP continues past the 4th season, I'll get the comfort to work on those fuzzy problems. But right now, I want to stick to items that are more clear.

GJP'S OWN FINDINGS

It does take a while for academic research to get published, but some of the results from the GJP has appeared in journals. The most recent journal article I can find is from January 2015:[6]

> This article extends psychological methods and concepts into a domain that is as profoundly consequential as it is poorly understood: intelligence analysis. We report findings from a geopolitical forecasting tournament that assessed the accuracy of more than 150,000 forecasts of 743 participants on 199 events occurring over two years. Participants were above average in intelligence and political knowledge relative to the general population. Individual differences in performance emerged, and forecasting skills were surprisingly consistent over time. Key predictors were: (a) dispositional variables of cognitive ability, political knowledge, and open-mindedness; (b) situational variables of training in probabilistic reasoning and participation in collaborative teams that shared information and discussed rationales (Mellers, Ungar, et al., 2014); and (c) behavioral variables of deliberation time and frequency of belief updating. We developed a profile of the best forecasters; they were better at inductive reasoning, pattern detection, cognitive flexibility, and open-mindedness. They had greater understanding of geopolitics, training in probabilistic reasoning, and opportunities to succeed in cognitively enriched team environments. Last but not least, they viewed forecasting as a skill that required deliberate practice, sustained effort, and constant monitoring of current affairs.

I think that abstract is accessible to the non-academic, but let's look at the media coverage of this:[7]

> "'Most people would expect to find domain experts doing well in their domain,' says Nick Hare, one of the super-forecasters (informally, they go by 'supers') whose performance in the project landed him an invitation to the Good Judgment Project's annual summer conference. But, in fact, 'there are people who are good at all domains'—outperforming even specialists. And they could hold the key to reconfiguring the way intelligence services think about making predictions in the future.

CONTINUED ON PAGE 22

WHAT I'VE LEARNED … | FROM PAGE 21

> "So, what makes Hare such a good forecaster? His success, he says, comes down not to knowledge but his capacity for 'active, open-minded thinking': applying the scientific method to look rigorously at data, rather than seeking to impose a given narrative on a situation."

I think this is really key. The point is to consider possibilities that might not accord with what you expect. In my own case, I'm looking at the feedback to try to improve, and I'm thinking of using these approaches in forecasts in my own job in insurance research.

Too often we may settle on an answer or forecast too quickly, based on our biases. The following:

- Actively seeking out information disconfirming our "gut instinct";
- Taking notes on our reasoning, to be referred to later;
- Regularly revisiting our predictions; and
- Conducting a post-mortem of the reasoning and process once an outcome is known;

are all great techniques I've learned (or re-learned, the hard way) by participating in the Good Judgment Project. I hope it continues for a fifth season, so I can continue to improve … and perhaps some of y'all will join me!

Mary Pat Campbell

Mary Pat Campbell, FSA, MAAA, PRM, is VP, Insurance Research at Conning in Hartford, Conn. She can be reached at marypat.campbell@gmail.com.

RELATED ARTICLES FROM SOA NEWSLETTERS

Wolzenski, Ben. "Predictably Irrational, by Dan Ariely The Hidden Forces that Shape our Decisions". Forecasting & Futurism Newsletter, December 2014. *https://soa.org/Library/Newsletters/Forecasting-Futurism/2014/december/ffn-2014-iss10.pdf*

Campbell, Mary Pat. "Know Thyself and Others". The Stepping Stone, May 2011. *https://www.soa.org/library/newsletters/stepping-stone/2011/may/stp-2011-iss42-campbell.aspx*

And maybe y'all can find more F&F articles that are related. ▼

ENDNOTES

1. Jay Ulfelder: Crowds Aren't Magic. May 20, 2014. http://goodjudgmentproject.com/blog/?p=215
2. About IARPA, accessed 2 March 2015. http://www.iarpa.gov/index.php/about-iarpa
3. "The Good Judgment Project: Improving intelligence estimates to support decision-makers", CHIPS Magazine, Jan – Mar 2015. http://www.doncio.navy.mil/chips/ArticleDetails.aspx?ID=5976
4. Spiegel, Alex. "So You Think You're Smarter than a CIA Agent" NPR parallels blog, April 2, 2014. http://www.npr.org/blogs/parallels/2014/04/02/297839429/-so-you-think-youre-smarter-than-a-cia-agent?utm_source=digg&utm_medium=email
5. Brier Score. Wikipedia. Accessed 2 March 2015. http://en.wikipedia.org/wiki/Brier_score
6. Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E., & Tetlock, P. (2015, January 12). The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics. Journal of Experimental Psychology: Applied. Advance online publication. http://dx.doi.org/10.1037/xap0000040
7. Burton, Tara Isabella. "Could you be a 'super-forecaster'?". BBC Future. 20 January 2015. http://www.bbc.com/future/story/20150120-are-you-a-super-forecaster

SVD of Weighted or Missing Data

By Brian Holland

Matrix decomposition techniques such as singular value decomposition (SVD), shed light on data's underlying structure. My article in the last newsletter describes SVD and described an example with unemployment rates. Unemployment rates were laid out in a grid, by county and month. The main features were quickly identified and then clustering techniques were applied among time periods—confirming the narrative that we already know about unemployment rates since 1990.

Insurance data, however, are rarely laid out in such a clean fashion. What happens when there are missing values? What happens when some values, like decrements, are based on little exposure and therefore volatile? These problems are not unique to insurance but actually quite widespread. Techniques are being developed and refined to address similar applications. This note gives a quick overview of approaches and points to further sources.

NONINSURANCE APPLICATIONS

There are a few incomplete data sets that we all encounter. Recommender systems are based on user ratings or purchases. None of us has watched—or at least rated—all the offerings on Netflix, with the possible exception of a few procrastinating actuarial students. The matrix of ratings by user and movie is mostly holes, i.e., it is a sparse matrix where most of the cells in any particular row or column are empty. So, how do you decompose that? For another example, images are matrices of pixels. What if many pixels are missing? How can objects be identified or images cleaned or completed? There is such a huge case for estimating who wants what, it is no wonder there are so many papers on the topic.

ACTUARIAL APPLICATIONS

There are several cases where clustering incomplete data would be highly useful. How many x-factor sets are needed? Or more generally, how many sets of assumptions are needed? The data are incomplete because the population is still running out. Much of actuarial work involves estimating the incomplete parts of a matrix: experience that has not yet evolved by policy duration or certain attained ages, for example. These matrix completion efforts could inform the actuary's decision on estimates for areas of thin exposure.

MISSING DATA: FILLING THE GAP

There are several approaches to estimating the missing data. The paper *Methods for Large Scale SVD with Missing Values* (Kurucz, Benczur, and Chalogany) focuses on estimating missing Netflix ratings. The dataset is much larger than actuaries are likely to encounter in practice. The authors conclude that a modified Lanczos algorithm is somewhat better than the power algorithm. The expectation maximization (EM) algorithm used proceeds with two steps after initializing the blanks: SVD is performed, blanks are re-estimated from the SVD results. The process repeats until adequate convergence.

Three initializations are tested for the missing values: either zeros, averages, or an item-based recommender using an adjusted cosine distance. Initialization with zeros or averages led to slow, if any, convergence. As for the algorithm itself, the authors used SVD as implemented by the Lanczos algorithm in *svdpack*, an R package, with some modifications. They also described and tested a modified power iteration method, but found that it over-fit the data.

Actuaries need to deal with weights in addition to missing data, and also include prior knowledge. In *Fast Regularized Low Rank Approximation of Weighted Data Sets*, Saptarshi Das and Arnold Neumaier present just such an approach. Regularization is applied to impose prior knowledge, in this case smoothness of a series, and also to avoid over-fitting. Imposing smoothness requirements is reminiscent of familiar actuarial techniques like graduation. The usual SVD formula is modified to apply a matrix of weights for the SVD error terms and also regularization terms for the left and right singular vectors. The example on which the method is tested is noise removal from astronomical images.

A more direct link to statistical models is shown in *Forecasting Time Series of Inhomogeneous Poisson Processes with Application to Call Center Workforce Management* by Haipeng Shen and Jianhua Huang. Calls into call centers are

CONTINUED ON **PAGE 24**

treated as a Poisson. They vary by time of day. The calls by time of day (bucketed) and day form a matrix, the expected value of which is the Poisson parameter. This parameter is transformed via a link function and decomposed to find latent parameters by maximizing likelihood. The Poisson factor model is fit to reduce the dimensions of the matrix. The ultimate goal of this paper is to fit a time series onto the factor score series. There are clear similarities to many insurance processes. The link function and probability-oriented approach could fit well within an actuarial context.

Brian Holland

Brian D. Holland, FSA, MAAA, is director and actuary, of Life and Individual A&H Experience Studies at AIG. He also serves as Vice-Chair of the Forecasting and Futurism Section Council. He can be reached at brian.holland@aig.com.

SVD's related technique, principal component analysis (PCA), is reformulated as a maximum likelihood solution of a latent variable model in *Pattern Recognition and Machine Learning* by Christopher Bishop. This reformulation is dubbed "probabilistic PCA." In this Bayesian treatment there are prior and posterior distributions of the latent principal components. Placing PCA in the familiar Bayesian framework brings all the familiar Bayesian advantages. For example, explicit expression of opinion and updating the model for new events.

All of these techniques hold promise for actuaries. Financial models and experience contain a vast array of decrements. Linking these decrements with dimension reduction techniques, and linking those further with Bayesian techniques, can yield a communicable overview of both existing models and also actual experience. ▼

Calibrating Risk Score Model with Partial Credibility

By Shea Parkes and Brad Armstrong

Risk adjustment models are commonly used in managed care programs to ensure that participating health plans are compensated based on their ability to manage costs, rather than on the underlying morbidity of their enrollees. The accuracy of the models can influence which plans receive a larger (or smaller) proportion of the funds.

A variety of claims-based risk adjustment models are available; each is designed to predict costs for a certain type of program, such as a Medicaid population versus a commercial population. However, the variety of managed care populations (and benefits) is much larger than the variety of off-the-shelf risk adjustment models that are available. It is inevitable that any specific program will exhibit characteristics—reimbursement, covered benefits, prevalence, and severity of disease states—that are different from those assumed by even the most appropriate model available. For example, a common concern in Medicaid is that reimbursement varies materially between states. The target program may have higher hospital reimbursement and lower professional reimbursement than other programs, or vice versa.

Although the off-the-shelf model may still do an acceptable job of predicting costs, it is likely that the accuracy of the model could be improved by recalibrating it to better fit the specific population for which it is being used. Most risk adjustment models are based on linear regression, so a common method of adjusting the model is to estimate new parameters (or weights) for the population of interest.

However, estimating new weights is only appropriate if the population is large enough to provide credible estimates of all the potential coefficients, especially those associated with less prevalent conditions or disease states. For example, the population may be large enough to support adjustments for more common conditions such as diabetes, but adjustments for less common conditions, such as tuberculosis or rare genetic conditions, may be based on a small sample of a few individuals and not fully credible. The off-the-shelf models represent valuable learnings from a very large, very credible data source. Instead of estimating completely new weights, it is possible to use a technique known as ridge regression to only adjust the coefficients that are credibly different for the target population. This can result in a model that is better than either of the off-the-shelf coefficients, or one that is completely retrained on the target population.

Definitions of "better" are often nebulous, especially when dealing with concurrent risk scores. In this case, "better" means that the model produces a lower error metric on a new dataset (other than that used to train it). If "better" were instead focused on the lowest error metric on the dataset used to train the model, then the fully re-estimated model will be best as long as it is optimizing the corresponding loss function. In the following example, the new dataset used to judge performance was claim experience from a different year of the program for the same population used to re-estimate the model.

AN APPLIED EXAMPLE

Recently, we have been exploring different techniques to recalibrate the Medicaid Rx (MRx) model to better fit specific populations. Medicaid Rx, a risk adjustment model designed for Medicaid populations, uses enrollment and National Drug Code (NDC) coded pharmacy claim data to assign individuals to age and gender categories and to flag each member for the presence of a variety of medical conditions, which are identified based on pharmacy utilization. The age/gender buckets and condition flags are then used as features in a linear regression model that predicts a risk score for each member. While we wanted to keep the variables used in the standard Medicaid Rx model intact, our goal was to reweight these variables in the linear model to better fit the characteristics of specific Medicaid programs and to improve the accuracy of the predictions on new data.

With enough data and experience, one way to accomplish this would be simply to take the known normalized costs of individuals, and fit a new linear regression model with the same features as the standard MRx model in order to completely recreate the coefficients from scratch. However, some populations are not large enough to be considered fully credible on their own. In this example, we focused on a population with approximately 30,000 members, which is not large enough to warrant full credibility. Instead of completely re-

CONTINUED ON PAGE 26

CALIBRATING RISK SCORE … | FROM PAGE 25

training the MRx model, we used the standard MRx weights as a starting point, but made adjustments to the coefficients of the model based on evidence from the population data. To strike this balance, we used a penalized regression and cross validation to choose a reasonable point between the standard weights and completely retrained weights.

Our linear regression model for creating new weights included all of the features of the standard MRx model (the demographic and condition variables), but also included an additional "offset" variable that represented the original model's risk score prediction. In a standard linear regression with conditional Gaussian response, this is equivalent to fitting a new model on the residuals of the original model. However, the "offset" paradigm can still apply in a generalized linear model setting.

Adding this new variable effectively meant that the coefficients estimated for all of the other features in the model could be interpreted as "deltas," or the adjustments that should be made to the standard/original weights. We then estimated the delta-coefficients with a ridge regression penalty, optimized via cross-fold validation. The ridge regression penalizes a model for the sum of its squared coefficients; this tends to prefer models with smaller coefficients versus those with wildly large coefficients even if the latter are slightly more accurate on the training data set. Because the coefficients were in fact "deltas" from the original coefficients, this essentially favors models that are closer to the off-the-shelf model. An alternative interpretation is that we put strong Bayesian priors on each coefficient, centered at the values used in the standard MRx model. Because the ridge regression framework adds a larger penalty as the coefficients for each variable get further away from zero, the tendency of the model was to use values close to the standard weights unless there was strong evidence in the population data that a certain coefficient should be changed. Even better, since the size of the ridge penalty can be scaled using a parameter, we are able to tune the procedure to vary the amount of credibility given to the population data.

Figure 1 shows how the values of the coefficients for select features change as different levels of credibility are given to target population data.

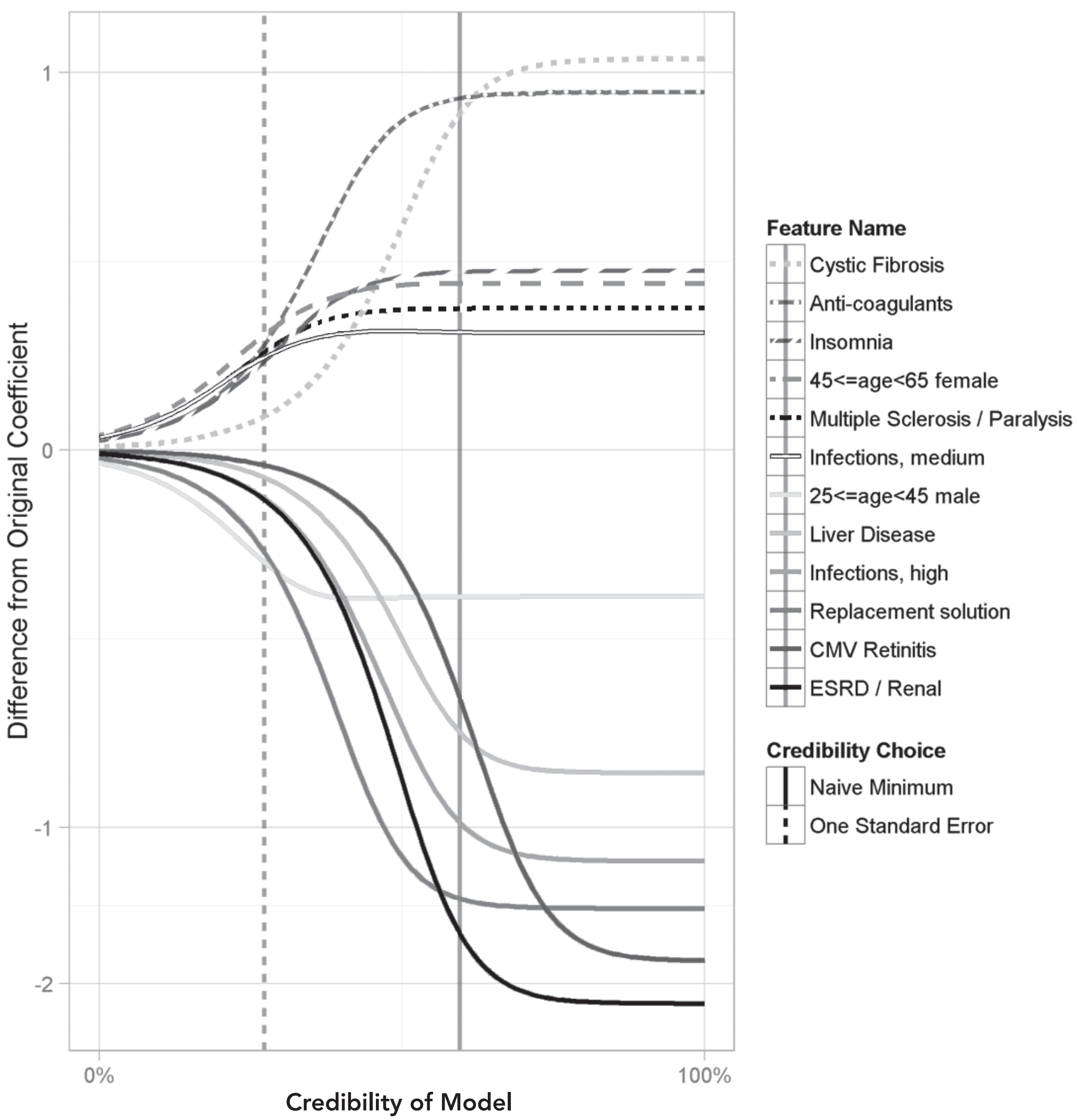
The outcome variable of our regression model was the normalized cost of each individual. The outcomes for very high cost individuals were capped at the 99.5th percentile cost of the population. This was done to avoid having a handful of observations inordinately affect the values of the coefficients estimated by the regression model. For example, one very high-cost individual flagged for a certain condition could singlehandedly push the coefficient associated with that condition much higher than it should be. By introducing the cap to the outcome variable, that individual would still be considered high-cost in the regression, but not by several orders of magnitude, which could swamp the importance of all other observations with that condition. This was especially important for the process of cross validation explained below.

To perform the ridge regression with cross validation, we used the glmnet package in R, which allows the user to fit a regression model with a ridge penalty, a lasso penalty, or a blend between the two (elastic net penalty). A lasso model penalizes the sum of the absolute values of the coefficients, while the ridge model penalizes the sum of the squared coefficients. By using the ridge penalty, the regression produced non-zero delta-coefficients for all of the features in the model, but the size of the adjustment varied based on the evidence in the population data. Using a lasso penalty would have made the delta-coefficient for many of the features zero, while only making adjustments to coefficients for which there was strong evidence. While the lasso approach could also produce reasonable results, we chose ridge regression based on a prior assumption that none of the coefficients were precisely centered for the target population.

CHOOSING A SPECIFIC SET OF COEFFICIENTS

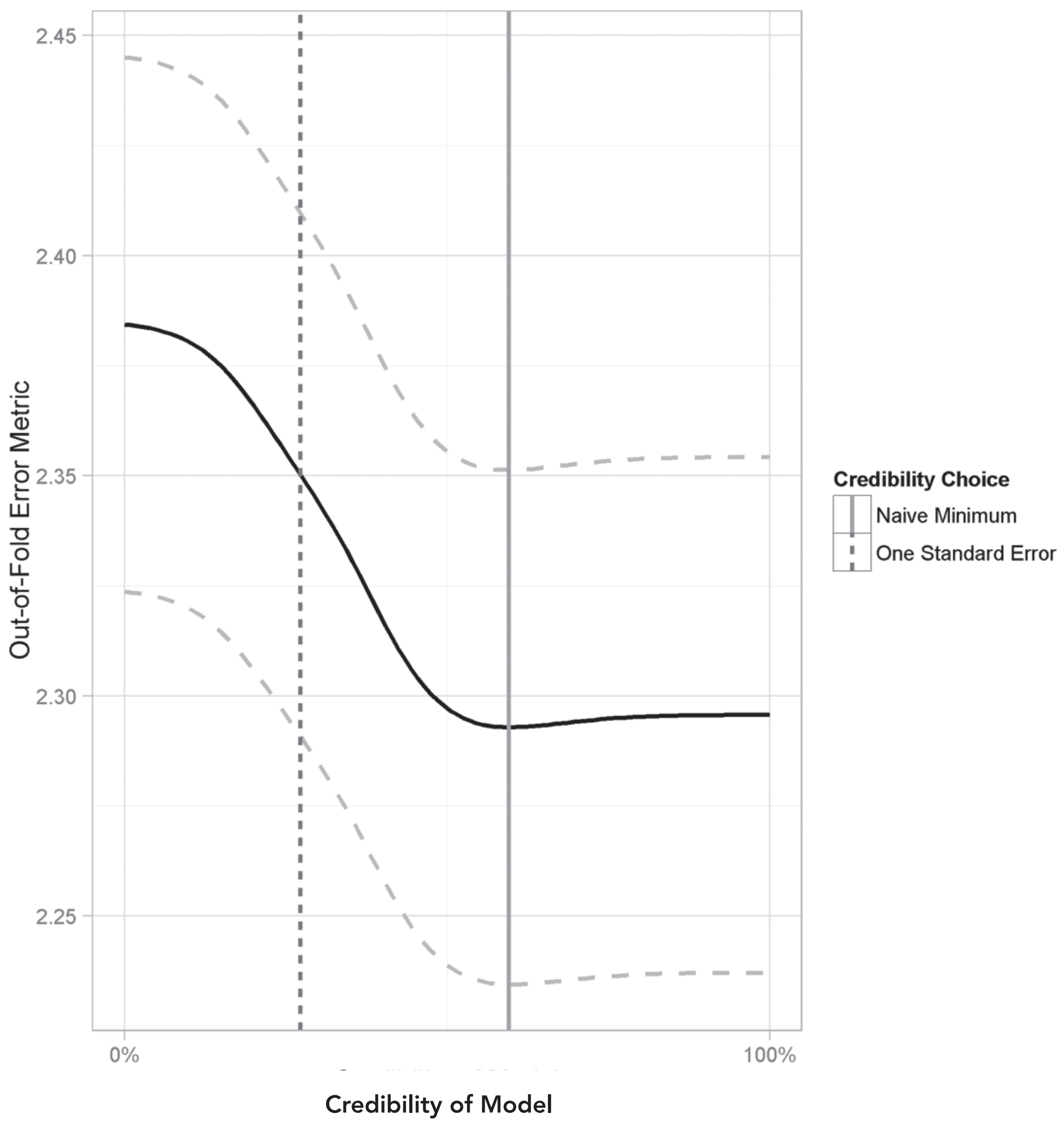
To decide how strong a ridge penalty to apply, we utilized 10-fold cross-validation within the training data. This means the training observations were divided into 10 segments, and the regression was performed 10 times, leaving a segment of the data out each time. For each fit, the model was judged against this smaller portion of the training data that was currently withheld, generating a cross-validated error metric. In theory, this produces a more realistic estimate of model performance on new data. There is still uncertainty about how new data might differ from the training data, so

Figure 1: Values of Coefficients for Select Features



CONTINUED ON **PAGE 28**

Figure 2: Error Estimates for Penalty Values



even this estimate of accuracy should be used with caution. For this application we utilized root-mean-square-error as the error metric, after capping extremely high cost members' outcomes to bound their influence. The insights should be the same for any reasonable choice of error metric.

This whole cross-validation procedure was repeated for different sizes of ridge penalty to produce a range of generalization error estimates for different penalty sizes. Instead of picking the penalty value with the absolute best cross-validated error estimate, we chose a slightly simpler (closer to off-the-shelf) model that was within one standard error of the minimum cross-validated error estimate. This is a standard convention to protect against overfitting, because resampling the training data does not truly reflect the new data to which one might want to generalize.

Figure 2 displays the error estimate for a range of penalty values. For our final model, we chose to use a penalty value for which the error estimate was within one standard error of the minimum, in order to prevent overfitting to new data.

In this example, our goal was to generalize to the next year of claims. Upon actual application, it was shown that the penalized model produced a better average error metric on the new year of data than the off-the-shelf model, and one very similar to the fully re-trained model. The specific error metrics are presented in the table below:

Model Description	Error Metric on Next Year of Claims
Off-The-Shelf	3.157
Partial Credibility	3.123
Fully Re-trained	3.123

While the penalized model exhibited the same level of predictive power as the fully re-trained model, the coefficients used in the penalized model appeared more reasonable and credible, because the weights for certain features were not based entirely on a low volume of observations. Using this methodology allowed us to still use the information contained in the standard weights of the MRx model, but to adjust them slightly to better accommodate the characteristics of this specific program. We recommend exploring this approach when trying to recalibrate a model for a population that is of a moderate size, but perhaps not fully credible. ▼

REFERENCES

1. Department of Family and Preventive Medicine, University of California, San Diego. Medicaid Rx. http://medicaidrx.ucsd.edu/

2. H. Wickham. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

3. Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL http://www.jstatsoft.org/v33/i01/.

4. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

5. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Science & Business Media, 2009.

Shea Parkes

Shea Parkes, FSA, MAAA, is an actuary at Milliman in Indianapolis. He can be reached at shea.parkes@milliman.com.

Brad Armstrong

Brad Armstrong, FSA, MAAA, is an associate actuary at Milliman in Indianapolis. He can be reached at brad armstrong@milliman.com.

How to Build a Model

By Richard Xu, Dihui Lai, Minyu Cao, Scott Rushing, and Tim Rozar

This article first appeared as appendix b of the Lapse Modeling for the Post-Level Period: A Practical Application of Predictive Modeling *report, sponsored by the SOA's Committee on Finance Research. It is reprinted here with permission.*

Building an effective and robust model requires a solid foundation in statistics and practical experience in statistical applications. For those wanting to increase their modeling skills, we recommend further study of statistical algorithms (such as GLM and decision trees) and additional development of applicable technical skills.

This Appendix serves as an introduction to a few basic modeling techniques. For a more complete and comprehensive understanding of statistical modeling, a formal study program would be beneficial.

The software and programming language used for this example is called R and is accessible to the public as an open-source application. There are no license restrictions. The system is expandable by design and offers very advanced graphic capabilities. As of June 2014, there are more than 5,800 add-on packages and more than 120,000 functions available under the R framework. R is developed based on a modern statistical language, which is very close to C/C++. A large online community is available to support learning, in addition to the built-in help system.

However, the learning curve for learning the R language and software environment can be quite steep. Additionally, there are limitations in using R such as the demands on memory, single thread in CPU utilization, limited graphic user interface, limited GUI, etc. Some of these problems can be addressed by the many add-on packages.

The example that follows is based on a hypothetical dataset and is intended for educational purposes. The data file is attached to this document and can be downloaded from SOA website where the main document is located. A few simple steps are provided to demonstrate a simplified approach to building a model in R.

Note: The commands that need to be entered into R are displayed in ***bold italics***, while the return from the R software is in `this Courier font.` Please note that R is a command-line system. To perform functions, a user is required to type in every command.

DATA LOADING

In the following R script, we assume the sample data file is called "SampleData2014SOAPM.csv", which is a comma delimited text file. To load the data into the R system, the following command should be executed, assuming the file is located in "C:/Data":

> lapseData <- read.csv("C:\\data\\SampleData-2014SOAPM.csv", header=TRUE)

The option of "header=TRUE" indicates that the names of the data fields are included in the data file. Since this is also the default setting, it can be ignored.

After reading the data, the R system assigns the whole dataset to an object called "lapseData". This object has the data structure called "data frame". The data frame structure is equivalent to a worksheet in an Excel file, with rows (record index) and columns (data fields) available for data manipulation.

R has other options to import data including from an Excel file, a database, the internet, or manually importing it into R by hard-coded R scripts.

DATA EXPLORATION

Once loaded, there are numerous ways to examine the data. Below are the two most common procedures to understanding the volume and characteristics of the data.

The 'summary' command returns the distribution of each field provided in the data.

>summary(lapseData)

```
 FaceAmount   PremiumMode RiskClass IssueAge  LapsedN           Exposure
 100-250K:28   Annual :70   NS:70   25-29:20  Min.   :    1.00 Min.   :    29.61
 250K-1M :28   monthly:70   SM:70   30-34:20  1st Qu.:   47.75 1st Qu.:   844.64
 50-100K :28                        35-39:20  Median :  417.00 Median :  7159.10
 GT1M    :28                        40-44:20  Mean   : 1735.03 Mean   : 24594.77
 LT50K   :28                        45-49:20  3rd Qu.: 1775.75 3rd Qu.: 24881.78
                                    50-54:20  Max.   :14712.00 Max.   :186853.50
                                    55-59:20
```

The 'head' command returns the first 6 records in "lapseData".

> head(lapseData)

```
   FaceAmount PremiumMode   RiskClass   IssueAge  LapsedN     Exposure
 1   100-250K     monthly          NS     25-29   1220        44507.43
 2   100-250K     monthly          NS     30-34   2023        65939.43
 3   100-250K     monthly          NS     35-39   2963        74532.25
 4   100-250K     monthly          NS     40-44   3779        75532.78
 5   100-250K     monthly          NS     45-49   4143        67085.31
 6   100-250K     monthly          NS     50-54   4267        59205.88
```

Other commands for data exploration include dim(), names(), tail(), aggregate() and many more.

MODEL CREATION

After the basic understanding of the data is obtained, one can start building a model. In the dataset, our target variable is the number of lapses per number of policies exposed per unit of time (in this case, one year).

In this sample model, the Poisson distribution is used and logarithm is the default link function.

The number of lapses is called 'LapsedN' in our model and 'Exposure' reflects the number of policies exposed for the corresponding duration. To reflect this in the model and since the link function is the logarithm, the offset is the logarithm of 'Exposure'.

$$\log(\text{LapsedN} / \text{Exposure}) = \log(\text{LapsedN}) - \log(\text{Exposure})$$

As we can see from the preceding equation, subtracting "log(Exposure)" on the right side of the equation as an offset is equivalent to dividing by 'Exposure' on the left side of the equation, which changes the lapse count to the lapse rate which is what is being modeled here.

> Model1 <- glm(LapsedN ~ offset(log(Exposure)) + FaceAmount + PremiumMode + RiskClass + IssueAge, family=poisson(), data=lapseData)

In the above command, "glm" is the specified model family, and 'family=poisson()' is the specified distribution. Since the default link function of logarithm is what's needed, it is not necessary to specify in the bracket. The target variable is 'LapsedN', and there are 4 explanatory variables: 'FaceAmount', 'PremiumMode', 'RiskClass', and 'IssueAge'.

After the model is fit with the data, the model results can be checked with the following command:

CONTINUED ON **PAGE 32**

HOW TO BUILD A MODEL | FROM PAGE 31

> summary(Model1)

```
Call:
glm(formula = LapsedN ~ offset(log(Exposure)) + FaceAmount + PremiumMode
             + RiskClass + IssueAge, family = poisson(), data = lapseData)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-14.4278   -1.7662   -0.1371    1.6875   14.4382

Coefficients:
                           Estimate    Std. Error    z value      Pr(>|z|)
(Intercept)               -2.872380   0.009227     -311.31       < 2e-16 ***
FaceAmount250K-1M          0.063109   0.004443       14.21       < 2e-16 ***
FaceAmount50-100K         -0.171759   0.010461      -16.42       < 2e-16 ***
FaceAmountGT1M             0.078315   0.007342       10.67       < 2e-16 ***
FaceAmountLT50K           -0.333405   0.054839       -6.08       1.2e-09 ***
PremiumModemonthly        -0.413123   0.004736      -87.23       < 2e-16 ***
RiskClassSM                0.061092   0.006221        9.82       < 2e-16 ***
IssueAge30-34              0.105852   0.010531       10.05       < 2e-16 ***
IssueAge35-39              0.207189   0.010053       20.61       < 2e-16 ***
IssueAge40-44              0.301690   0.009946       30.33       < 2e-16 ***
IssueAge45-49              0.398574   0.009955       40.04       < 2e-16 ***
IssueAge50-54              0.474820   0.010132       46.86       < 2e-16 ***
IssueAge55-59              0.537070   0.010541       50.95       < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 (Dispersion parameter for poisson family taken to be 1)

Null deviance: 18197.4  on 139  degrees of freedom
Residual deviance:  2395.1  on 127  degrees of freedom
AIC: 3461.5
```

The distribution of deviance residuals is displayed in a summary format. The deviance residuals are similar to the standardized error terms.

Following the list of deviance residuals are the predictor variable list, the coefficients and other statistics which have the same format as a standard Ordinary Least Squares (OLS) model.

The deviances of a null model and the current model are stated at the end of the output. The AIC (Akaike information criterion) is also calculated for generic GLM distributions such as the Poisson, Gamma, and Normal distributions. The last line of the output displays the number of iterations of numeric analysis in the GLM algorithm.

After initial iterations of the model, higher orders of covariates and cross-terms need to be considered to account for the significant interactive effects between the predictor variables.

For this particular sample dataset, the cross term between 'PremiumMode' and 'IssueAge' can be tested to improve the model's predictive power.

Here are the R script and results:

> Model2 <- glm(LapsedN~offset(log(Exposure))+FaceAmount+PremiumMode+RiskClass + IssueAge + PremiumMode:IssueAge, family=poisson(),data=lapseData)
> summary(Model2)

```
Call:
glm(formula = LapsedN ~ offset(log(Exposure)) + FaceAmount + PremiumMode
             + RiskClass + IssueAge + PremiumMode:IssueAge, family = poisson(),
             data = lapseData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.9645  -1.3702  -0.0883   1.0014   5.2205

Coefficients:
                                 Estimate    Std. Error  z value   Pr(>|z|)
(Intercept)                     -2.772184   0.010354  -267.74    < 2e-16  ***
FaceAmount250K-1M                0.063053   0.004444   14.188    < 2e-16  ***
FaceAmount50-100K               -0.182915   0.010468  -17.474    < 2e-16  ***
FaceAmountGT1M                   0.083356   0.007345   11.348    < 2e-16  ***
FaceAmountLT50K                 -0.341975   0.054840   -6.236    4.49e-10 ***
PremiumModemonthly              -0.775494   0.020480  -37.866    < 2e-16  ***
RiskClassSM                      0.060230   0.006220    9.684    < 2e-16  ***
IssueAge30-34                    0.079421   0.012037    6.598    4.17e-11 ***
IssueAge35-39                    0.149634   0.011510   13.001    < 2e-16  ***
IssueAge40-44                    0.206212   0.011421   18.056    < 2e-16  ***
IssueAge45-49                    0.272848   0.011431   23.870    < 2e-16  ***
IssueAge50-54                    0.321941   0.011663   27.602    < 2e-16  ***
IssueAge55-59                    0.362104   0.012186   29.716    < 2e-16  ***
PremiumModemonthly:IssueAge30-34  0.067422   0.024827    2.716    0.00661  **
PremiumModemonthly:IssueAge35-39  0.188513   0.023577    7.996    1.29e-15 ***
PremiumModemonthly:IssueAge40-44  0.343423   0.023177   14.817    < 2e-16  ***
PremiumModemonthly:IssueAge45-49  0.468724   0.023182   20.219    < 2e-16  ***
PremiumModemonthly:IssueAge50-54  0.578425   0.023481   24.634    < 2e-16  ***
PremiumModemonthly:IssueAge55-59  0.659804   0.024234   27.227    < 2e-16  ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 18197.4  on 139  degrees of freedom
Residual deviance:   427.6  on 121  degrees of freedom
AIC: 1506

Number of Fisher Scoring iterations: 4
```

CONTINUED ON **PAGE 34**

As seen in the result, by adding the cross term, the AIC is significantly reduced from 3462 to 1506 and residual deviance decreases from 2,395 to 428. The inclusion of the cross term substantially improves our model's performance.

It is tempting to add as many cross-terms as possible to improve the model performance. However, it is important to balance the model fit with both simplicity and business judgment.

A model should be validated to test its effectiveness. There are many techniques available for this purpose; however, they will not be discussed here due to the scope of this brief introduction.

PREDICTION AND RESULT VISUALIZATION

After the model is built, the model is then used to predict lapse rates.

```
> lapseData$pred <-predict(Model1, lapseData,
type="response")
```

In this command, the model "Model1" is applied to the dataset "lapseData". The prediction is the response of the model, which is the predicted mean value. Other options are available, such as confidence level and uncertainty.

With both predicted values and observed values available, plots can be made to illustrate the model's goodness of fit by comparing the model's predicted lapses to the actual lapses.

R has very strong built-in graphic capabilities. There are numerous packages available for data visualization. It is simple to export the plots to the clipboard or a stand-alone file in popular formats such as .pdf or .bmp. To make an A/E plot, data needs to be calculated and aggregated. In the following example, A/E is calculated by premium mode and risk class.

```
> byPred <- aggregate(pred ~
PremiumMode+RiskClass, data = lapseData, FUN =
sum)
> byObsv <- aggregate(LapsedN ~
PremiumMode+RiskClass, data = lapseData, FUN =
sum)
> AERatio <- byObsv[,3]/byPred[,3]
> AERatio
```

```
[1] 1.0030546 0.9903241 0.9767918 1.0517889
```

The last command displays the values of A/E ratios. Once the ratios are calculated, the following R scripts will plot the ratio, display the title, show the label on the X-axis, and draw a red line at 100% as reference:

```
> plot(AERatio,xlab="PremiumMode+RiskClass",
ylab="AE Ratio", xaxt='n', ylim=c(0.9,1.1), pch=18)
> title("A/E vs. Premium Mode and Risk Class")
> axis(1, at=1:4,labels=c("NS-Annual","NS-
Monthly","SM-Anual","SM-Monthly"), las=0)
> abline(1,0,col="red")
```

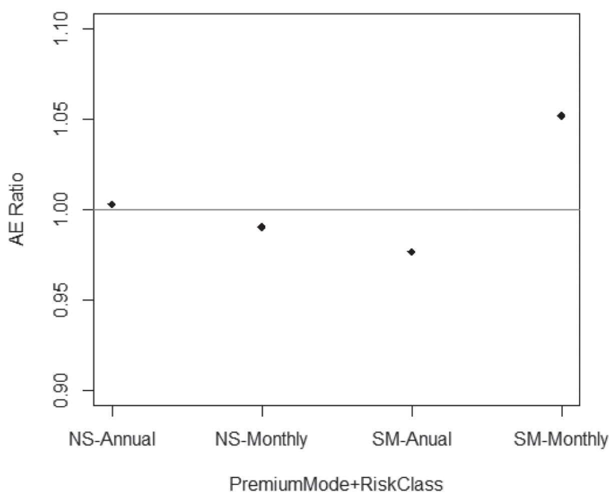
Richard Xu

Richard Xu, FSA, Ph.D., is senior data scientist at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *rxu@rgare.com.*

Dihui Lai

Dihui Lai, Ph.D., is data scientist analyst at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *dlai@rgare.com.*

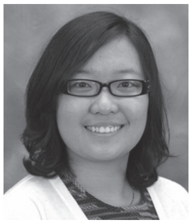
A/E vs. Premium Mode and Risk Class



Another option is to export the results data to a file and perform data visualization in other applications such as Excel. This approach is probably more appealing to actuaries since actuaries are more familiar with Excel. The following script can be used to accomplish this:

> write.csv(lapseData,"modelDataFile.csv")

With this command, R will write the contents of "lapseData" into a file in the default directory with the name "modelDataFile.csv." ▼



Minyu Cao

Minyu Cao, ASA, is an assistant actuary at RGA Reinsurance Company in Chesterfield, Mo. She can be reached at mcao@rgare.com.

Scott Rushing

Scott Rushing, FSA, is VP & actuary, Head of Global Research at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at srushing@rgare.com.

Tim Rozar

Tim Rozar, FSA, CERA, is senior vice president at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at trozar@rgare.com

Simple Rating Systems:
Entry-level sports forecasting

By Doug Norris

As actuaries, we typically focus our predictive efforts in a relatively small niche area—for instance, I primarily focus on commercial health care pricing, reserving, and strategy. However, I would speculate that most of us learned our love for mathematics and forecasting long before we were formally trained in actuarial techniques.

Growing up in the suburbs of Seattle, I was a sports fan. In particular, I was fascinated by sports statistics. I would invent baseball games using my card collection and a set of oddball dice (for anyone looking to follow in my footsteps, 10-sided dice are incredibly handy). I pored over Bill James's annual *Baseball Abstract* editions. I tracked the statistics for my Little League team. I played APBA[1], Strat-O-Matic, and SherCo simulation games (and still play in a Strat-O-Matic hockey league to this day[2]). In 1994, I started one of the first sports websites, The Goaltender Home Page,[3] dedicated to preserving the history and numbers of hockey's unsung heroes.

A seminal moment in my actuarial career came with Bill James's 1985 *Baseball Abstract*, where James develops a rudimentary predictive model called "Brock-2." Given a baseball player's statistics to date, this model attempted to "complete" the player's remaining career. I dutifully reproduced the formulas in my parents' Apple II+ (fortunately, we had the model with 64k of memory, which was almost enough to reproduce the model), and thus began my first foray into predictive modeling. As an 11-year-old, I had a very powerful thought—how cool would it be if we could predict everything in sports? (I now realize that not only are sports inherently not perfectly predictable, but that those unpredictable aspects are the things that make sports the most fun.)

One of the most basic elements of sports forecasting involves predicting the winner of an upcoming game. Many are interested in being able to do this well, with billions of dollars bet on just the most recent Super Bowl alone.[4] This article describes a simple method for forecasting sports outcomes; in fact, the name itself has an appealing simplicity to it. However, the method is flexible enough to incorporate personal touches and improvements based on your own experience and judgment.

WHAT IS A SIMPLE RATING SYSTEM?

Nearly everyone who attempts to predict the outcome of sporting events realizes that past performance is a key indicator of future success. For instance, when an undefeated team plays a winless team, the undefeated team usually wins.

Predictions incorporating each team's point totals involve a trade-off—instead of focusing on what we are truly interested in (wins), we emphasize a proximate measure (points are not identical to wins, but points represent a "currency" that is used to purchase wins). Therefore, although a team's overall success is intimately intertwined with how well they produce points and prevent their opponents from producing points, counting points (instead of counting wins) results in a loss of specificity. However, this loss is offset by a gain in data—although most sporting events produce only one winner and one loser, each event produces many more points (or goals, or runs, or whatever translation your sport of choice uses). The increase in data helps to offset small sample size variation to some degree, and the trade-off typically results in increased predictive ability.[5] The Simple Rating System (SRS) method incorporates point totals, but takes things one step further.

Consider a six-team hockey league, with franchises named the Alligators, the Badgers, the Conquistadors, the Dragons, the Eagles, and the Falcons. So far in the season, each team has played three games, as shown in Figure 1.

Figure 1: League Outcomes to Date

Road Team	Goals	Home Team	Goals
Alligators	2	Badgers	5
Conquistadors	1	Dragons	0
Eagles	2	Falcons	2
Badgers	3	Alligators	0
Dragons	1	Falcons	4
Eagles	1	Conquistadors	3
Badgers	5	Dragons	1
Eagles	0	Alligators	3
Conquistadors	1	Falcons	1

The Falcons will next visit the Alligators, the Dragons will visit the Eagles, and the Badgers will visit the Conquistadors. Our goal is to provide our best estimate of who will win each game (and by how much). A good first step would be based on how each team has performed so far, so let's look at that to guide us, in Figure 2.

Figure 2: League Performance to Date

	Games	Wins	Losses	Ties	Win %	Goals Scored	Opponent Goals	Avg. Margin Victory
Badgers	3	3	0	0	100%	13	3	+3.33
Conquistadors	3	2	0	1	83%	5	2	+1.00
Falcons	3	1	0	2	67%	7	4	+1.00
Alligators	3	1	2	0	33%	5	8	-1.00
Eagles	3	0	2	1	17%	3	8	-1.67
Dragons	3	0	3	0	0%	2	10	-2.67

As you can see, the Badgers are playing very well, and the Dragons are playing very poorly, with the other teams spread out in between. The Falcons are outscoring their opponents by one goal per game, while the Alligators are being outscored by one goal per game, so we might reasonably predict that the Falcons will beat the Alligators by two goals in their next contest. Similarly, we might predict the Eagles (-1.67 goals/game) to defeat the Dragons (-2.67 goals/game) by one goal, and the Badgers (+3.33 goals/game) to outscore the Conquistadors (+1.0 goal/game) by 2.33 goals.

But wait—the Badgers have played a pretty weak schedule thus far (facing the Alligators twice and the Dragons once). Could their observed dominance be merely a reflection of their strength of schedule, and not their true ability? The Badgers' typical opponent has lost games by an average of 1.56 goals, so if playing against a truly "average" opponent, we would expect the Badgers to win by (+3.33 goals) + (-1.56 goals) = +1.78 goals. Let's revisit all six teams, focusing on their average margins of victory along with their strength of schedule (opponents' average margins of victory), as shown in Figure 3.

Figure 3: Simple Rating System: First Iteration

	Avg. Margin	Schedule Strength	Adj. Avg. Margin
Badgers	3.33	-1.56	1.78
Conquistadors	1.00	-1.11	-0.11
Falcons	1.00	-1.11	-0.11
Alligators	-1.00	1.67	0.67
Eagles	-1.67	0.33	-1.33
Dragons	-2.67	1.78	-0.89

Note: Adjusted average margin of victory = average margin of victory + schedule strength)

CONTINUED ON **PAGE 38**

We can see that the teams with worse records have generally played a stronger schedule, facing stronger opponents (and vice versa)—this makes sense intuitively for two reasons: first, we are measuring average margins of victory, and teams with losing records necessarily have given their opponents more wins than losses. Second, teams with losing records don't get to play themselves. Typically, these disparities are stronger when teams have only played a few games (and have a disproportionate share of their games against one team).

Revisiting our upcoming games, and noting that the adjusted average margin of victory (AAMV) represents how a team might fare against an "average" opponent, we would adjust our predictions such that the Alligators (+0.67 AAMV) would be favored by 0.78 goals over the Falcons (-0.11 AAMV), the Dragons (-0.89 AAMV) would be favored by 0.44 goals over the Eagles (-1.33 AAMV), and the Badgers (+1.78 AAMV) would be favored by 1.89 goals over the Conquistadors (-0.11 AAMV). Note that the predicted outcomes of our three games have changed considerably (with the overall winner changing in two of the three predictions).

At this point, you may be wondering—if we believe that the AAMV values represent a more accurate "team strength" metric, why aren't we using them to determine each team's schedule strength? Yes, we should be using the AAMV to develop an updated strength of schedule (SOS) estimate for each team, which in turn produces an improved estimate of AAMV (and so forth). In the end, we're looking for AAMV estimates that, when used to compute schedule strength estimates, produce the same AAMV estimates in return. In linear algebra parlance:

$AAMV_O$ = initial average margin of victory for each team
SOS_n = average $AAMV_n$ of each team's opponents (weighted by times played)
$AAMV_1 = AAMV_O + SOS_O$
$AAMV_{n+1} = AAMV_O + SOS_n$

We would like to find values for $AAMV_n$ such that $AAMV_n$ equals $AAMV_{n+1}$. If S (short for "schedule") represents the matrix where $S_{x,y}$ counts the proportion of times that team *x* has played team *y*, we know that

$$AAMV_{n+1} = AAMV_O + S * AAMV_n$$

For $AAMV_n$ to equal $AAMV_{n+1}$, we must satisfy:

Solving for $AAMV_n$:

$$AAMV_n - S * AAMV_n = AAMV_O$$
$$(I - S) * AAMV_n = AAMV_O$$
$$AAMV_n = (I - S)^{-1} * AAMV_O$$

Where *I* is the $n_x n$ identity matrix. Those of us who have taken linear algebra are happy to see the end point; however, in this case, the (I – S) matrix proves to be singular (and therefore non-invertible).[6] However, we can solve the problem numerically, and compare the differences of successive iterations; our hope is that the sum of the absolute value of these differences becomes sufficiently small after a large number of iterations, in which case we have found a convergent solution.[7] For our mythical hockey league, Figure 4 shows the unique convergent solution (and final Simple Rating System margins of victory for each team).

Ultimately, our SRS algorithm predicts the Alligators (+0.55 SRS) to be favored by 0.72 goals over the Falcons (-0.17

Figure 4: Simple Rating System: First Iteration

	Avg. Margin Victory	Avg. Opp. SRS Margin	SRS Margin
Badgers	3.33	-0.21	3.12
Conquistadors	1.00	-1.17	-0.17
Falcons	1.00	-1.17	-0.17
Alligators	-1.00	1.55	0.55
Eagles	-1.67	0.07	-1.60
Dragons	-2.67	0.93	-1.74

SRS), the Eagles (-1.60 SRS) to be favored by 0.14 goals over the Dragons (-1.74 SRS), and the Badgers (+3.12 SRS) to be favored by 3.29 goals over the Conquistadors (-0.17 SRS).

HOW CAN WE IMPROVE UPON THE SIMPLE RATING SYSTEM?

First and foremost, the SRS algorithm is not guaranteed to converge, particularly when the network of games played is sparse. For instance, when each team has only played one game, then an infinite number of convergent solutions exist. Related to this, until the SRS algorithm has enough data to work with, the credibility of the predictions suffers. Similar to pricing an insurance product, sports forecasters will typically blend experience data with a "manual rate" until the experience data can stand on its own legs. This manual rate could be based upon prior years' data, or built using other information (there are some interesting agent-based approaches to this), and then massaged by knowledge of the participants.

Speaking of credibility, one flaw of the SRS algorithm (as presented here) is that it considers all data fed into it to be of equal credibility. In reality, a team with sufficient sample size is more likely to perform at the level of its recent performance than at the level of earlier events. It is a simple matter to tweak the SRS algorithm to allow for different outcome weights (as for what those weights should be, that's where art meets science).

Similarly, there are many things that are "known" about sports. First, teams typically perform better in their home environment (this probably makes sense intuitively, even if you aren't a sports fan). Second, outlier performances, where one team dominates an opponent to an excessive degree—such as a football game with a score of 55-0, or a baseball game with a score of 14-1—can have a disproportionate effect on SRS algorithms, because in games where the outcome is decided early, teams do not necessarily finish the game at their "true" ability level. Third, team composition can change throughout the course of a season, which is due to trades, promotions and demotions, coaching changes, injuries, and other factors (which can also affect individual, i.e. non-team, athletes, such as tennis players and golfers). All of these can be accounted for, using judgment and experience, in SRS algorithms. One additional modification (that shows some predictive "lift") considers offensive and defensive contributions separately, as (for instance) a team that scores proficiently against a good defense might deserve more credit than would be expected by comparing against the opponent's overall AAMV.

Sports fans reading this article are probably already thinking of additional improvements that could be made to the SRS algorithm, including sport-specific nuances to improve the predictive nature of the methodology. Of course, this is the fun of predictive models, and in sports forecasting, the truly brilliant modifications are proprietary and confidential. With that said, if you come up with anything compelling, I'd love to hear more about your efforts. Remember that illegal gambling is illegal (hence the term "illegal gambling"), and that this article is for entertainment purposes only. ▼

ENDNOTES

1 A game company once named American Professional Baseball Association.

2 See the National Strat-O-Matic Hockey League at http://nshl.org.

3 See http://hockeygoalies.org. Clearly, I have invested more in data improvements than in aesthetics.

4 American Gaming Association (January 22, 2015). Illegal Super Bowl bets to total $3.8 billion this year. Retrieved March 27, 2015, from http://www.americangaming.org/newsroom/press-releases/illegal-super-bowl-bets-to-total-38-billion-this-year.

5 Baseball-Reference.com (February 20, 2015). Pythagorean Theorem of Baseball. Retrieved March 27, 2015, from http://www.baseball-reference.com/bullpen/Pythagorean_Theorem_of_Baseball.

6 This is a rather fun proof left for the reader. First, prove that each row of (I – S) sums to zero. What does this imply about the triangularized matrix?

7 In this case, all of the linear algebra holds up to (but not including) the matrix inversion step, meaning that the solution (if it exists) is not necessarily unique.

Doug Norris

Doug Norris, FSA, MAAA, Ph.D., is a principal and consulting actuary at Milliman in Denver, Colo. He can be reached at doug.norris@milliman.com.

Stepping Out

By Bryon Robidoux

An overview of the Predictive Analytics World (PAW) conference.

At the end of March, I attended the PAW conference in San Francisco. I first discuss the general structure of the conference so you can relate it to the SOA conferences. Then I will discuss the R class that I attended. Lastly, I will explain the many important overtones that rang through the different presentations at the conference.

The conference was setup in three parts: pre-conference, conference, and post-conference. The pre- and post-conference were day-long classes. They had classes for beginners all the way to experts. The conference itself was only two days long. It was divided into two tracks. The first track was for anyone. The second track was for practitioners and experts. I mainly stayed with the first track. For each day of the conference, there was a breakfast keynote address and a lunch keynote address and then after each keynote we would break off into our track of interest. I really enjoyed having classes before and after the conference. Given that I am only allowed one paid conference per year, this allowed me to attend a couple of training courses in addition to going to the conference. If the Society is not already doing this, I think this would a great idea to try.

The *R for Predictive Modeling: A Hands-On Introduction* class was by far my favorite part of the conference. I easily could have attended five days of this class and skipped the rest of the conference. It was supposed to be a hands-on class, but there was so much information that there was no way to listen to him and have time to run the code myself. Max Kuhn coauthored a book *Applied Predictive Modeling*, which I highly recommend. This class was really just a cliff note version of his book. In this class, he focused a lot on preparing data features for the models, such as centering, scaling, and removing skew using the Box-Cox transformation. He discussed which models require the transformations, such as linear regression or principal component analysis, versus models that do not need the data transformation, such as decision trees. He focused on preparing testing and training data. It is obvious that the less data the actuary has the more important it is to "spend" the data wisely. He explained the importance of stratified sampling to make sure the training and test sets have the same distribution of event outcomes. Given the correct features are in the model, he discussed the difference between bootstrapping and K-fold cross-validation of the model. Bootstrapping is sampling with replacement and the sample is the same size as the original dataset. K-fold cross-validation is sampling a subset of the test data and using the remaining portion to test the prediction. This is done K times, where K is usually five or 10. He talked at great length about comparing different model measures, such as R squared, root mean square error. His preference was to use mean square error over R squared because you can get a high R squared by just having a very volatile distribution. He talked at great length about comparing different classification models using Lift Curves and the Receiver Operating Curves (ROC). Overall I would give this class an A. If you can't make it to the Advanced Business Analytics Seminar in Philadelphia this year, then this would be a great alternative. As an added bonus, it is given four or five times a year. I hope to attend one of Max's classes next year.

In terms of the conference, it was very interesting because I was able to see all the different uses for predictive modeling outside of an insurance context. I really enjoyed seeing how other disciplines frame their problems and derive solutions. This usually gives a fresh insight into problems that I am trying to solve. There definitely were some interesting themes throughout the conference. In the keynote address, "The Revolution in Retail Customer Intelligence," the major theme was to look past the demographic information and collect or transform the data in a form that has a better "signal" of intent and behavior of the individual. His example was trying to explain why customers behave as they do when shopping on the Internet and how can you turn this knowledge of behavior into more sales. For an actuarial example of a GMWB policy, the age of an annuitant really gives little information about the potential behavior of an annuitant, but knowing the number of years until retirement or number of alternative income sources gives far more color on utilization of the benefits. If the data is encoded in such a way that the modeler can determine intent and behavior, the predictive model will be better at prediction.

The second keynote, "New Challenges: From Predictive to Prescriptive to Automated Analytics" was trying to address the difference between data versus information. Data can seem infinite and it is currently growing at an exponential rate, but the information extracted from the data is finite. The issue is how are actuaries, statisticians, and data scientists going to be able to process all of the data to extract the information. The short answer is that they are not. Dell's answer to the solution is to automate model calibration, modeling, and responses to the output. In the future, the models will self-calibrate and have little or no interface. In the future, the data scientist will be mostly involved in making sure the process is behaving as expected. Personally, I would not want to be a data scientist at that point because the creativity and art would be sucked out of the science.

The third presentation I want to mention is "Making Impacts Through Analytics," by Bin Mu from MetLife. In many presentations including this one, presenters discussed the importance of communication. If the modeler does not have the ability to communicate the importance of the model and its potential return on investment, then the model is of little use. I especially enjoyed this presentation because Bin took this concept a step further and described a framework for modeling: define business objectives, analyze information and draw insights, identify actions from insights, and measure business impact. At the end of the day, if you are able to describe your projects with these four steps, then you will always be able to justify your work to the C-Suite. The point he drove home over and over again was to design experiments so that the results and outcomes are easily measurable. One of his major frustrations was with internal clients asking him to do studies, but the business owner really had no plans for using the information outside of a standard report that will more than likely be ignored. He turned the question around and asked the client how to make a request actionable and how can the request improve the business.

I REALLY ENJOYED SEEING HOW OTHER DISCIPLINES FRAME THEIR PROBLEMS AND DERIVE SOLUTIONS. THIS USUALLY GIVES A FREASH INSIGHT INTO PROBLEMS THAT I AM TRYING TO SOLVE.

The last item that I would like to mention is about the expert panel on "Education and Training Options for Predictive Analytics." In this panel, they debated the lack of data scientists in the future and the education required in getting the industry up to speed. It was explained that there will be a deficit of 140,000 to 190,000 jobs in data science in the next three years. Schools are scrambling to train people for these positions. They discussed at length the steps of evolving into the field:

1. Read high-level books;
2. Watch instructions on YouTube, which I thought was interesting;
3. Massively Open Online Classes (MOOC), for example Coursera.org;
4. Certification from a university to supplement your professional experience; and
5. Get a Master's Degree.

The last topic discussed was the qualities of a good data scientist, which is very similar to the qualities of an actuary. Besides being good at math, a person needs to be creative, innovative, and most importantly, have good communication skills.

I do think it would be an interesting debate to determine when someone becomes a data scientist versus a statistician versus an actuary. How and where do you draw the line between these disciplines? The bigger question that went through my mind as I was enjoying the conference was, as

CONTINUED ON **PAGE 42**

STEPPING OUT | FROM PAGE 41

Bryon Robidoux

actuaries, how do we fit into the big data/predictive modeling revolution? Are we on the sidelines or are we active participants? Does the combination of course P and course C qualify us as a data scientist? It was interesting, while talking to a gentleman in my R class, he explained that he wanted to use predictive modeling to determine when Pepsi-Cola machines were most likely to fail. I have 25,000 good friends that can help him solve that problem! The disturbing part of the conference for me was that I think I was the only actuary that attended. (They provided a mobile app, which allowed me to see the occupation of other participants at the conference.) Notice that the room full of statisticians and data scientists did not put on their list of possibilities of becoming a data scientist to first become an actuary. I would love to see how actuaries were perceived by this audience and the role we should play.

In conclusion, I was glad that I attended this conference. I definitely think it as worth my time to go, especially because of the day-long training classes before and after the conference. If possible, I would like to go again next year and take more advanced classes. (It doesn't hurt that I got to enjoy fresh crab on the Fisherman's Wharf while gazing at the Golden Gate Bridge in the background.) But the question that churns in mind is how do actuaries fit into the predictive analytics revolution? ▼

Bryon Robidoux, FSA, CERA, MAAA, is director of Hedging, Global Financial Solutions, at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at brobidoux@rgare.com.

What Big Data is, and How to Deal with It

By Jeff Heaton

Big data is a frequent participant in headlines today. The amount of electronic data available is growing at an exponential rate. Every few years new terms such as megabyte, gigabyte, terabyte, petabyte, exabyte, and even zettabyte enter everyday vocabulary as a new measure for "really large data." Some estimates state that the total amount of electronic data by 2020 will exceed 35 zetabytes. But what exactly is "big data," and how much of those 35 zetabytes will a typical company actually need to process?

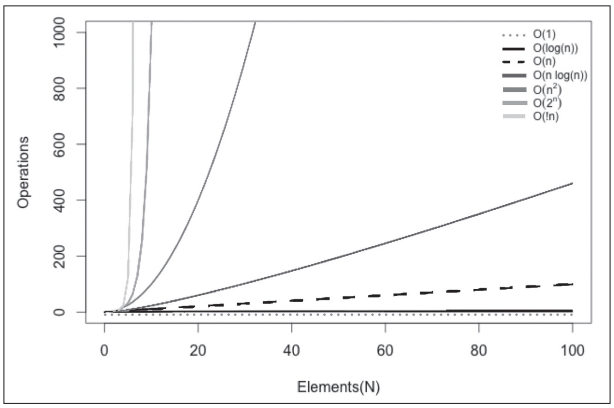
Much hype surrounds the term "big data," and several definitions exist for the term. One of the most useful definitions of big data comes from Wikipedia, "big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate." There are several thresholds effectively established by this definition. Will the data fit into a server's RAM? Will the data fit onto a single hard disk drive? As the size of the data grows more traditional tools begin to fail. There are a multitude of companies ready to sell you new tools to handle big data. Often these tools cost big dollars.

Going by these definitions, big data is nothing new. If your computer has 16K of RAM, then 17K is "big data." Back in the 1990s I had to make many modifications to a C++ application to allow it to make use of its full 2 MB of RAM. The Intel architecture of the time could only access 1MB of RAM at a time. My program had to share the lower 640K with DOS and map sections of the EMS memory into the upper 384 MB of the address space. Was this "big data?" In a sense it was "big data," the problem had become large enough that it no longer fit into RAM.

WHY IS BIG DATA HARD

Big data is hard because computer programs do not always scale well. In computer science, the scalability of a computer program is measured in something called big O notation. You may have heard of algorithms referred to as running in O(log N), O(N2) squared or even O(N!) exponential time. These refer to how well the program scales to its data set.

The most efficient computer program would be O(1) time. Such a program will always run in the same amount of time, regardless of how large the data set is. Consider a program that finds the first name in a list. Such a program will always take the same amount of time because it does not matter if the list has 10 items or 10 million items. Very few things run in O(1) time, however, O(n) is reasonably good as well. Consider if I asked you to find the longest name in a list. For this you must visit each item in the list, so it is O(n). Assuming *n* is the number of items in your list. The processing time should scale linearly. If it takes 10 minutes to process 10 items, it should take 100 minutes to process 100 if you are dealing with an O(N) algorithm.



Not every algorithm behaves linearly. Knowing the O-magnitude of an algorithm can help you decide which to use. The seven most common magnitudes are shown on the following chart.

As can be seen from the chart algorithms, the most favorable magnitude algorithms are O(1) , O(log(n)) and O(n). The least favorable are O(n2), O(2n) and o(!n).

If n is relatively small, it does not matter what the magnitude of your algorithm is. However, as n grows, so does the processing time of the algorithm. Some algorithms simply do not work with big data because of their magnitude. When dealing with a high-magnitude algorithm, and big data, it is often necessary to accept an approximation, rather than process the entire data set. Some algorithms that initially seem high-magnitude can be rewritten to be more efficient.

CONTINUED ON **PAGE 44**

WHAT BIG DATA IS … | FROM PAGE 43

THE LANGUAGE OF BIG DATA

Big data has its own terminology, just like any other field. Doug Laney, of the META Group defined datasets in terms of three V's. This has come to be known as the three V's of big data. The first V, volume, describes the size of the data set. This is the characteristic that frequently comes to mind when discussing big data. The second V, variety, describes the complexity of the data. When dealing with big data there will often be several large datasets of different variety. This can pose unique challenges for the algorithms that must process these datasets. The final V, velocity, describes the rate at which the data is changing. The underlying dataset will often change during the time that the big data algorithm is processing.

Velocity introduces streaming, which is another important big data concept. Streaming, or real-time processing, refers to a large amount of data that arrives continuously over time. The amount of data arriving in the stream may increase and decrease as the stream of information flows into your program. Examples of stream data include trading, fraud detection, system monitoring, and others.

Out-of-core, or external memory algorithms, is another important concept for big data. Such algorithms do not use computer RAM to process their datasets. It is very common practice to load an entire dataset into memory and then process it. However, this is not always necessary. Even if a low-magnitude O(N) algorithm is chosen, it will fail as soon as n grows to the point that the list can no longer fit into memory. Consider calculating the mean of numbers in a very large list. A computer program could read the list, number-by-number, and maintain two variables. The first variable keeps a sum of the numbers encountered, and the second variable keeps a count of the number elements processed so far. At the end, these two variables will hold the sum of the list, as well as the count of items in the list. Simply divide the sum by the count and you have the mean. It does not matter how large the list is, you will have sufficient memory to calculate this mean.

Vowpal Wabbit is a popular out-of-core machine learning framework. By using memory only as a cache, Vowpal Wabbit is capable of processing any size dataset. It might take Vowpal Wabbit a very long time to process a dataset; however, it would not run out of memory and crash, like many similar programs. This is very similar to how programs were written in the past when RAM was scarce. Modern computers, with their large memory systems, often encourage programmers to not pay attention to their memory usage. Programs that naively load entire datasets into RAM simply will not scale to large amounts of data.

TOOLS FOR BIG DATA

Two of the most commonly used tools for big data are Hadoop and Spark. The Apache Foundation manages both of these programs. Hadoop is the foundation upon which Spark is built. Hadoop provides distributed file storage and the communication infrastructure needed by Spark. Hadoop uses the map-reduce algorithm to perform distributed processing. Map-reduce requires considerable disk I/O, as large problem spaces are mapped into parts, and those parts combine and reduce into the ultimate solution. Spark uses Resilient Distributed Data (RDD) to break the problem into many pieces that can be processed in RAM on the nodes. Whereas Hadoop needs fast disk I/O, Spark needs considerable RAM. For the right tasks, this can mean processing time increases of 100 times compared to Hadoop alone.

One type of problem that excels under Spark is machine learning. The ability to break the problem into many units executed in RAM is very conducive to many machine learning algorithms. Spark has a model called MLlib, or Machine Learning Library that provides many machine learning models right out of the box. Hadoop, along with another Apache framework called Pig, is very good at performing traditional SQL queries over very large datasets.

TOWARD SCALABLE ALGORITHMS

Traditional programming wisdom says to first focus on getting a working program and optimize later. Donald Knuth is quoted as saying, "Premature optimization is the root of all evil (or at least most of it) in programming." While this is generally true, big data forces optimization to increase in priority. Analytics often forces many runs before the desired result is achieved. The shorter a runtime that you achieve, the more experimentation you can do.

Many common programming tasks have both naive and optimized implementations. Consider some of the following operations on a big list of numbers:

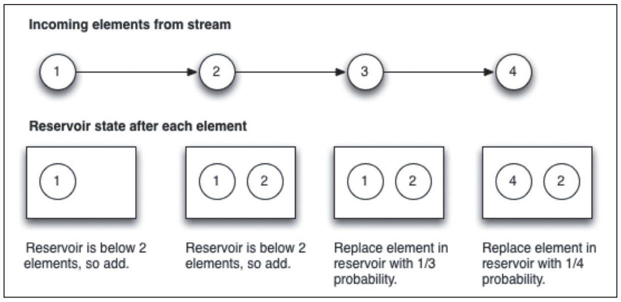
- Percentile and Quintile Estimation,
- Randomly sampling a subset,
- Sorting,
- Taking the mean,
- Taking the standard deviation,
- and more.

Each of the above algorithms has naïve and optimized approaches. Searching and sorting are among the most researched algorithms in computer science for efficiency. Consider the standard deviation, which normally requires two passes over the data. First you calculate the mean, and then you calculate the mean deviation of each data point from that mean. A naïve standard deviation calculation requires two passes over the data. There are algorithms that can do it with one pass. These same algorithms are also good for calculating the mean and standard deviation over an endless stream of numbers.

Reservoir sampling is a very common big data technique that can be used to randomly sample a set of numbers from a very large pool. Consider if you wanted to randomly choose two people from the world population. The naïve approach would be to visit each person in the world once to obtain an accurate count and place him or her into a consistent ordering. You would then select two random numbers up to the world population count. Using this number, you would now visit everyone in the world again, and stop at the index numbers that you randomly chose in the previous step.

This approach has several issues. First, between the two passes, people would have been born and died. Barring a large-scale natural disaster, the world population count would be higher for your second pass than the first. The sample would no longer be uniform, and would bias against those that were born since the first pass. This is the velocity problem of big data. However, the biggest problem is that it is potentially necessary to visit everyone twice. The two-pass method also becomes nearly impossible to use when dealing with an endless stream of data.

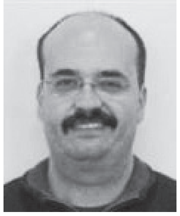
The following figure illustrates how to use reservoir sampling with a stream of numbers.



To sample two elements from a large stream of numbers you simply add the first two to the reservoir. When selecting the third element you now replace an element in the reservoir with #3 with a 1/3 probability. Likewise, for the fourth element you replace an element in the reservoir with a 1/4 probability. This continues for as much data as you have.

CONCLUSIONS

Big data presents many challenges for analytics systems. It is very important to choose tools and underlying algorithms that will scale to the size of your data. Data have a tendency to grow as systems mature. The sooner in the development cycle that you make scalability decisions the better. Tools designed to work with big data can help to facilitate this growth, even if you are not dealing with big data today. ▼



Jeff Heaton

Jeff Heaton is data scientist, Global R&D at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at jheaton@rgare.com.

A 'Hot Date' with *Julia*: Parallel Computations of Stochastic Valuations

By Charles Tsai

Meet "Julia," a free programming language licensed by MIT that may help you with parallel computing. It may be an alternative tool for those who are interested in nested stochastic processes for actuarial research (if not for regulatory compliance).

Nested stochastic processes may become more relevant and prevalent as stakeholders consider a broader spectrum of possible outcomes. Such "stochastic-in-stochastic" analyses often add color to actuaries' palette of tail risks and conditional tail dependencies (if any). However, they also introduce issues of runtime and memory allocation. The article "Nested Stochastic Pricing"[1] provides a comprehensive summary of nested stochastic applications in response to recent regulatory reforms. IFRS seems to require a comprehensive range of scenarios that reflects the full range of possible outcomes for calculating fulfillment cash flows. Economic capital calculations may likewise require stochastic-in-stochastic simulations. A practice that may have been previously deemed as a costly bonus may evolve into a minimum expectation for actuaries in the near future.

Nested stochastic processes may become more acceptable with parallel computations. One may boil down "parallel computing" to daily applications with an analogy. Imagine an investment banker who is planning a date with a lady. He barely has enough time to smoke, and he has completing the following four tasks in mind: 1) dress up, 2) buy flowers, 3) research a restaurant's menu, and 4) fold a thousand origami cranes. He has made these preparations in solo for all of his previous dates. Would it not be nice for him to have friends help him perform the latter three tasks *simultaneously*? Delegation may take some time, but it may be more efficient than performing all four tasks in sequence. Parallel computing is a form of dividing and conquering problems using multiple processes concurrently. It may help actuaries slam-dunk tasks like traversing a thousand scenarios, even if the tasks already take less time than folding a thousand origami cranes.

Julia allows users to distribute and execute processes (such as nested stochastic valuations) in parallel. In essence, a computer may have four Central Processing Units (CPUs) in resemblance to a soccer team with four members. Programmers can leverage Julia's multiprocessing environment to specify certain tasks to those CPUs on the bench. On the one hand, the art of scheduling may be a bulk process for infrequent and smaller tasks. On the other hand, the flexibility to pass messages to multiple processors may be one's niche in strategic scalability and performance. Actuaries may then manage disparate layers of stochastic simulations via a multiprocessing environment. Shorter runtimes may be a doomsday for a few students who use waiting time as an opportunity for studying. However, such efficiency opens doors to comprehensive iterations and widens windows of perspectives.

IS JULIA A DISRUPTIVE INNOVATION?

Julia has several features[2] that supplement its power in parallelism and distributed computation. Some features are for specialists like Sheldon Cooper (of *The Big Bang Theory*) while others may be easier for amateurs like me to appreciate.

- First, it is free and open sourced as licensed by MIT. Actuaries can share research results seamlessly at SOA/CAS events without worrying about whether the audiences have access to the same tools to review (and build upon) the findings.

- Second, users can define composite types that are equivalent to "objects" in other languages. These user-defined types can run "as fast and compact as built-ins".[3]

- Third, users can call C functions directly, and their programs' performances can approach those of languages like C. Such speed makes it a considerable alternative to proprietary computational software tools.[4]

- Fourth, one does not need to be a genius like Gaston Julia in order to learn the language. Justin Domke's blog post "Julia, Matlab, and C"[5] presents a crystal clear comparison of syntactic and runtime complexity tradeoffs. Learning Julia is like learning Matlab® and C++ for Towers Watson MoSes® simultaneously.

- Last but not least, Julia is a functional programming language like OCaml, which is adopted by niche firms like Jane Street. Functional programming frameworks can help actuaries adapt to and master recursions.

Julia also has several Achilles' heels that may significantly jeopardize its adoption among actuaries.

- One obstacle is communication. Due diligence may be lost in translation. A few know how to use and interpret proprietary actuarial software products due to limited availability. Fewer know how to read and review (or even find) its generated C++ codes. In a like manner, few have learned (or are willing to learn) the Julia language, and its graphical features are still under development. Some actuaries may still prefer parallel computations via multiple Microsoft Excel® sessions. Calibrations of Julia programs with validated Microsoft Excel® workbook models might just have exceeded paychecks.
- Another hindrance is the language's relative immaturity. Development commenced in 2009.[6] Its scale of recognition seems to be light years from the tipping point for a stabilized discussion ecosystem to exist. Online inquiries for relevant debugging notes make passing bills during gridlocks look easy. A tool may only be as valuable as its received appreciation.
- Lastly, the manipulation of processes in parallel computations requires an acute awareness of read-write conflicts. In light of the previous analogy, the banker may wish to match his suit with the flowers purchased, or the flowers purchased with the restaurant's cuisine. Tasks may not be completely independent from each other. Inexperienced users may inadvertently manipulate and designate processes in manners inconsistent with intentions.

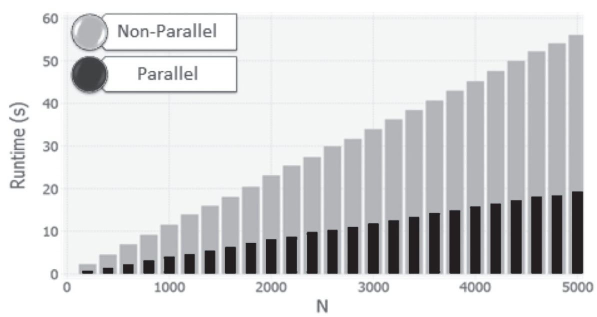
A SIMPLIFIED GMMB CASE STUDY

I have drafted an exemplary Julia application of an actuarial model. It is available at *https://github.com/Chuckles2013/GMMB_RSLN2_Julia*, and is an independent project for educational purposes only. All parameters and values have been arbitrarily chosen. The case study involves calculating the present values of liabilities for an extremely simplified Guaranteed Minimum Maturity Benefit (GMMB).

The scale of the project can be partitioned into two major layers. The first layer involves simulating parameters for N world scenarios. For simplicity, I have structured all key parameters to be the same across all N world scenarios. It is easy to see that one can simply modify the codes to utilize simulated parameter inputs for considering different world scenarios and economic environments. The second layer involves simulating fund returns for 1000 funds, from which one can derive a conditional tail expectation of liabilities. Both layers provide N figures of conditional tail expectations, from which one can extract a maximum level.

The superimposed bar graph below compares runtimes for non-parallel versus parallel computations under various numbers (*N*) of world scenarios. Four processors performed the parallel computations. The absolute values of the excess time elapsed are evident in the divergent gap.

Runtime Comparisons Across Different Numbers of Scenarios (When 4 CPUs Are Available)



CONTINUED ON **PAGE 48**

A '*HOT DATE*' WITH JULIA … | FROM PAGE 47

NEXT STEPS

One's vision for Julia in actuarial science can be the development of packages. A few companies were bold enough to have utilized R, and none has adopted (or even plan to leverage) Julia to my knowledge. Full adoption of Julia among actuaries within the next decade may be more of a fantasy than a reality, just as few actuaries have learned Python since its inception in 1991.[7] Nevertheless, open-source packages for broader usage are lower hanging fruit for intrigued actuaries to consider. To the best of my knowledge, there are no Julia packages similar to the lifecontingencies and actuar packages in R libraries. Templates of actuarial functions in Julia may capture more attention and appreciation for the beauty of parallel computations for nested stochastic valuations. ▼

Charles Tsai

Charles Tsai, ASA, is a Life Actuarial Analyst at AIG in Shanghai, China. He can be reached at *charles-cw.tsai@aig.com.*

ENDNOTES

1 "Nested Stochastic Pricing: The Time Has Come" by Milliman®'s Craig Reynolds and Sai Man is available at http://www.milliman.com/insight/insurance/pdfs/Nested-stochastic-pricing-The-time-has-come/

2 http://julialang.org/

3 http://nbviewer.ipython.org/github/bensadeghi/julia-datascience-talk/blob/master/datascience-talk.ipynb

4 Professor Fernández-Villaverde's "A Comparison of Programming Languages in Economics", which is available at www.econ.upenn.edu/~jesusfv/comparison_languages.pdf

5 http://justindomke.wordpress.com/2012/09/17/julia-matlab-and-c/

6 web.maths.unsw.edu.au/~mclean/talks/Julia_talk.pdf

5 This is a rather fun proof left for the reader. First, prove that each row of (I – S) sums to zero. What does this imply about the triangularized matrix?

7 http://svn.python.org/view/*checkout*/python/trunk/Misc/HISTORY

Forecasting & Futurism's Article Anthology

By Dave Snell

Issue	Date	Page	Author	SOA Member	Article	Comments
1	9/2009	3	Dave Snell	Y	Forecasting & Futurism Newsletter —A New Name and a New Dimension for Our Section	overview of the issue, with editorial on reason for the new name of the section
		5	Ben Wolzenski, Alan Mills	Y,Y	Introducing the New Forecasting and Futurism Professional Interest Section	new mission of the new section
		6	Alan Mills	Y	Introduction to Forecasting Methods for Actuaries	comparison of various forecasting methods (including predictive modeling)
		10	Scott McInturff	Y	The Delphi Method	detailed description of Delphi method with an actual case study
		17	Alan Mills	Y	White, Gray and Black Swans-Identifying, Forecasting and Managing Medical Expenditure Trend Drivers in a Complex World	tail events: identifying and predicting them
		22	Dave Snell	Y	Fortune's Formula: The Untold Story of the Scientific Betting System That Beat the Casinos and Wall Street—by William Poundstone	dangers of implicit belief in mathematical models
		23	Alan Mills	Y	Should Actuaries Get Another Job? - Nassim Taleb's Work and Its Significance for Actuaries	why forecasts fail, and how to avoid the classic mistakes
2	7/2010[1]	3	Dave Snell	Y	Judgmental Methods, Collaboration, Contests and More!	overview of the issue, and editorial on judgmental forecasting methods
		5	Alan Mills	Y	Want to Win an iPad?	forecasting contest
		6	Alan Mills	Y	Best Methods and Practices in Judgmental Forecasting	summary of types of bias, and comparison of various judgmental forecasting results
		14	Ben Wolzenski	Y	Future Opportunities in the Life Insurance Industry—Views of a Delphi Study	summary of the extensive (250+ page) F&F Blue Ocean Delphi study
		19	Scott McInturff	Y	The Wisdom of Crowds by James Surowiecki	review of book explaining that crowds can often give more accurate predictions than experts
		23	Dave Snell	Y	Complexity: A Guided Tour by Melanie Mitchell, Ph.D.	review of intro book on complexity sciences
		25	Mike Lindstrom	Y	Forecasting Judgment: The Netflix Prize and Collaborative Filtering	discussion of models to predict preferences of movie rental customers

CONTINUED ON **PAGE 50**

Issue	Date	Page	Author	SOA Member	Article	Comments
		27	John L. Petersen	N[2]	A New End, A New Beginning - Preparing for Life as We Don't Know It	a futurists view of how we need to change our thinking to embrace the future
		32	Ben Wadsley	Y	Living with Actuarial Black Swans—a Discussion with Nassim Nicholas Taleb	Taleb's views of the futility of predictive models
3	7/2011	3	Dave Snell	Y	Look! Up in the Sky! It's Super Actuary!	issue overview and an editorial on the need for cross-functional perspectives
		5	Ben Wadsley	Y	Moving the Ball Forward	advantages of volunteering
		6	Ben Wadsley	Y	Are Genetic Algorithms Even Applicable to Actuaries?	genetic algorithms used to determine optimal bond portfolio mix
		12	Min Deng	Y	Complexity Science Enters the Actuarial Classrooms	description of complexity science classes at Maryville University
		14	Doug King	Y	Judgmental Forecasting in Determining Policyholder Behavior Assumptions	judgmental forecasting and dynamic lapse assumptions
		16	Frank Grossman	Y	An Alternate View of Future Mortality	reasons why mortality may not continue to improve
		18	Scott McInturff	Y	The Perfect Swarm: The Science of Complexity in Everyday Life By Len Fisher	book review of book on swarm intelligence (locusts, ants, bees)
		22	Dave Snell	Y	Complexity Sciences—Simplified!	short descriptions of deterministic chaos, behavioral economics, predictive modeling, fractal geometry, genetic algorithms, network theory, cellular automata
4	1/2012	3	Dave Snell	Y	Actuaries: Do We Know Our Limits?	how to recognize and extend our limits in modeling and forecasting
		5	Ben Wadsley	Y	Volunteerism is Rewarding!	benefits of volunteering
		6	Donald Krouse	Y	Exploring, Growing, Learning …	basic applications of complexity
		7	Ben Wolzenski	Y	Delphi Studies Past, Present and Future	predictions from an SOA F&F Delphi study from 2004 are compared with amazing results
		12	Charles Brass	N[2]	Investigating the Future: Lessons from the "Scene of the Crime"	how a futurist is like a crime scene investigator (CSI)
		16	Frank Grossman	Y	Hidden in Plain Sight	anamorphoses - looking at problems from different perspectives
		18	Dave Snell	Y	Standing Room Only! Complexity Grows at Annual Meeting	four sessions on complexity sciences get packed attendance
		21	Ben Wadsley	Y	F&F 2nd Annual iPad 2 Forecasting Competition	contest competition problem and rules - civilian unemployment rate

Issue	Date	Page	Author	SOA Member	Article	Comments
		23	Dave Snell	Y	When Algebra Gets Chaotic	how deterministic chaos can surface in a seemingly simple equation
		27	Brian Grossmiller	Y	Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life by Albert-Laszlo Barabasi	book review on network theory
5	7/2012	3	Dave Snell	Y	"What's in a Name?"	new names gaining popularity: complexity sciences, predictive modeling, etc.
		5	Donald Krouse	Y	Future = Unknown = Risk = Opportunity	thoughts from Life & Annuity Symposium
		6	Brian Grossmiller	Y	Artificial Intelligence: What Is It and How Can I Use It?	key definitions of AI and a source for more education
		8	Richard Xu	Y	How to Win an iPad2	refresher on regression and time series models
		12	Donald Krouse	Y	Challenging Old Paradigms—What Are You Going to Do?	retirement schemes - how to plan for your retirement
		16	Kurt Wrobel	Y	The Actuarial Profession and Complex Models: Knowing the Limits of Our Knowledge	changes in the business environment have impacted our ability to ensure that our organizations make appropriate decisions
		20	Ben Wolzenski	Y	Growing Artificial Societies: Social Science from the Bottom Up, by Joshua M. Epstein and Robert Axtell	a description of Sugarscape - one of the first examples of active agent modeling
		21	Dave Snell	Y	Bad Science, by Ben Goldacre	a good primer on how clinical studies should and should not be conducted; and on how statistics are used and misused to manipulate public opinion
6	12/2012	3	Dave Snell	Y	Sugar and Spice, and Everything Nice!	issue overview and link to the infamous Target predictive modeling debacle
		5	Clark Ramsey	Y	The Times They Are A-Changin'	summary of F&F accomplishments during the year
		7	Dave Snell	Y	Genetic Algorithms—Useful, Fun and Easy!	genetic algorithms with minimal genetics
		17	Min Deng, Guangwei Fan	Y	Actuarial Communication at a University	written and oral communication classes for actuaries
		19	Brian Grossmiller	Y	Futurism in the Workplace	futurism tools: Delphi, Cross-Impact Analysis, Decision Modeling, Environmental Scanning, Futures Wheel, Gaming and Simulation, Relevance Trees, Scenarios, System Dynamics, Trend Impact Analysis, Visioning

CONTINUED ON **PAGE 52**

Issue	Date	Page	Author	SOA Member	Article	Comments
		25	Ben Wolzenski	Y	Artificial Society Modeling with Sugarscape	Insurance application of Sugarscape, from Growing Artificial Societies
		28	Mark S. Dion	N[3]	Predictive Modeling, A Life underwriter's Primer	Predictive Modeling from the perspective of a life underwriter
		36	Glenda Maki	N[4]	What Is Complexity Science?	Interview with six complexity science pioneers
7	7/2013	3	Dave Snell	Y	"If I only had a brain"	issue overview and sources for education on artificial intelligence
		6	Jeff Heaton	N[5]	Bayesian Networks for Predictive Modeling	an introduction to Bayesian Networks and how to train them for predictive modeling
		12	Richard Xu	Y	Predictive Modeling	Predictive Modeling: basics of PM and types of models in common use
		18	Brian Grossmiller, Doug Norris	Y,Y	Hidden Markov Models and You	putting Hidden Markov Models to work solving insurance problems
		24	Clark Ramsey	Y	Dark Side of the Moon	essay on how the new complexity science tools are like being able to finally see the dark side of the moon
		26	Scott McInturff	Y	The Signal and the Noise: Why So Many Predictions Fail—but Some Don't - by Nate Silver	review of book on the techniques of Nate Silver, a forecaster of amazing accuracy
		28	Ben Wolzenski	Y	Delphi Study 2000—Predictions for 2010 and 2050	Delphi study where the prediction for total life insurance in force was correct (in trillions of dollars) to two decimal places!
		30	Alberto Abalo	Y	Forecasting & Futurism Third Annual iPad Contest: Build a Genetic Algorithm	rules for the F&F genetic algorithm contest
		32	Jon Deuchler	Y	Global Trends 2030: Alternative Worlds - Report by the National Intelligence Council summary	several scenarios of the future, from a coalition of 17 U.S. government agencies
		33	David Wheeler	N[6]	Behavioral Economics: Implications for Actuarial Science and Enterprise Risk Management	how the human mind can be tricked into jumping to CONCLUSIONS vs GQNSLHSIQNS
		35	Mike Lindstrom	Y	The 10th Speculative Fiction Contest Forecasting and Futurism Section Prize Winner - The Weight Of Certainty—Selected Stories Of Steve Mathys	fiction from the winner of the actuarial speculative fiction contest
		36	Darrick Fulton	N[7]	Spreadsheet Controls ... How to Prevent a Fire	A CPA and Auditor discusses prudent controls to manage spreadsheets
		38	Dave Snell	Y	I Held a Human Brain!	educational opportunities for actuaries - both distance learning and very hands on local opportunities
8	12/2013	3	Dave Snell	Y	Embrace the Future—But Beware the Smug	essay on the need to keep the classical tool set even though the new complexity tools seem shiny and great

Issue	Date	Page	Author	SOA Member	Article	Comments
		7	Alberto Abalo	Y	The Future Ain't What It Used to Be	what newborn son can teach us about the need for insurance and proper risk management
		9	Ben Wolzenski	Y	A Return Visit to the Sugarscape	When you give the model a mind of its own, you might be surprised by its thoughts!
		12	Brian Grossmiller, Doug Norris	Y,Y	Hidden Markov Models and You, Part Two	a really deep dive into the intricacies of Hidden Markov Models
		24	Jeff Heaton	N[5]	A NEAT Approach to Neural Network Structure	Neuroevolution of Augmenting Topologies (NEAT) - what it is and how it is useful
		30	Richard Xu	Y	Predictive Modeling in Insurance: Modeling Process	flow diagram of the predictive modeling process … and an example application
		35	Ben Wolzenski	Y	"Land This Plane"—A Delphi Study about Long-Term Care in the United States	Only 10% of Americans 50+ have LTC coverage - Delphi study to find solutions
		40	Paula Hodges	Y	Delphi Study in Real Time—Life and Annuity Products and Product Development	Delphi study with new twist: instant feedback and quick path to next round
		41	Dave Snell	Y	Genetic Algorithms Revisited—A Simplification and a Free Tool for Excel Users	article describes a free Excel workbook that gives you a genetic algorithm shell
		46	Alberto Abalo, Doug Norris	Y	And the Winner Is ...	Jeff Heaton wins iPad for best genetic algorithm (next article)
		47	Jeff Heaton	N[5]	Diagnosing Breast Tumor Malignancy with a Genetic Algorithm and RBF Network	uses genetic algorithm to devise a radial basis function and predict malignancy of breast tumors
		51	Steve Epner	N[8]	Are Spreadsheets Sabotaging Your Accuracy?	spreadsheets may be everywhere; but we must learn to control and manage them.
9	7/2014	3	Dave Snell	Y	If More Precision Is Always The Answer, Have We Forgotten The Question?	issue overview and discussion of precision vs. accuracy
		7	Alberto Abalo	Y	What's In A Name?	discussion of the section name and what it means
		8	Geof Hileman	Y	Roughly Right	discussion of the dangers of precision
		10	Charles Brass	N[2]	The Past Is No More Certain Than The Future - Decision-Making In The Face Of Unavoidable Uncertainty	one crime, two unanimous jury decisions - except they don't agree!
		14	Doug Norris	Y	Forecasting & Futurism Fourth Annual Contest: How do YOU Forecast?	contest rules for the 2014 forecasting contest

CONTINUED ON **PAGE 54**

ARTICLE ANTHOLOGY … | FROM PAGE 53

Issue	Date	Page	Author	SOA Member	Article	Comments
		17	Jeff Heaton	N[5]	An Introduction to Deep Learning	intro to the machine learning approach made famous by IBM's Watson
		22	Richard Xu, Dihui Lai	Y, N[9]	Big Data in Life Insurance: Does it exist? If so, how should we handle it?	ways to deal with the mounting challenges of capacity and speed as data scales up rapidly in size.
		26	Ben Wolzenski, Ron Hagelman	Y, N[10]	A Conversation About the Delphi Study on Long-Term Care Financing Solutions	discussion of a collaboration with LTC section and non-actuaries
		28	Dave Snell	Y	Warm and Fuzzy ... And Real!	part one of a primer on Fuzzy (versus classical Boolean, or 'crisp') logic
		35	Jeff Heaton	N[5]	Fuzzy Logic in R	example of how fuzzy logic can be utilized in R
10	12/2014	3	Dave Snell	Y	Master of Accurate Calculations … Really?	issue overview and editorial on the need to embrace new forecasting tools or risk obsolescence
		6	Doug Norris	Y	Five Years Is A Lifetime - (Personal Forecasting)	summary of the rapid rate of changes impacting actuaries and their opportunities
		8	Ben Wolzenski	Y	Predictably Irrational, by Dan Ariely - The Hidden Forces that Shape our Decisions	review of best selling Behavioral Economics book, with chapter-by-chapter synopses
		12	Kurt Wrobel	Y	Risk Management and the Power of Simplicity	reprint from *The Actuary* on dangers of overreliance on models - particularly complex models
		15	Brian Holland	Y	Unsupervised Methods: An Overview for Actuaries	application of dimension reduction techniques to better visualize data clusters and facilitate predictive models
		22	Richard Xu, Dihui Lai	Y, N[9]	Predictive Modeling Series: Data Clustering and Its Application in Insurance	introduction to predictive modeling clustering procedures and the associated feature extraction, proximity measures, and risk seqmentation
		26	Geof Hileman, Claire Bobst	Y, N[11]	A Nearest Neighbors Approach To Risk Adjustment	using R to perform k-nearest neighbors analysis for health risk adjustment models
		31	Jeff Heaton	N[5]	Agent Based Modeling With RePast Py	how to create a simple agent based model with the Recursive Porous Agent Simulation Toolkit (RePast) for Python
		36	Jeff Heaton	N[5]	Modeling With Python and Scikit-Learn	ScikitLearn examples of basic linear regression, decision trees generation, and visualization
		41	Doug Norris	Y	Parables And Prophecies Prevent Proper Predictive Prowess (human biases in forecasting)	description of human biases and how they influence of predictions … also mitigation ideas
		44	Dave Snell	Y	Warm and Fuzzy ... and Real! – Part 2	second part (with underwriting examples and R code) of primer on Fuzzy Logic
		51	F&F Council	Y	Forecasting & Futurism 4th Annual Contest	rules and criteria for the F&F Predictions and Forecasting contest

Issue	Date	Page	Author	SOA Member	Article	Comments
11	7/2015	3	Dave Snell	Y	Forecasting & Futurism – "One of the Best Kept Secrets of the SOA"	issue introduction with summaries of all articles - theme is behavioral economics
		7	Doug Norris	Y	Can't Win for Losing	chairperson article - describes winner's curse
		9	Ben Wolzenski	Y	Review: Why Smart People Make Big Money Mistakes and How to Correct Them: Lessons from the New Science of Behavioral Economics, by Gary Belsky and Thomas Gilovich	book review of behavioral economics book
		13	Tyson Mohr	Y	Review: Thinking Fast & Slow, by Daniel Kahneman	book review of behavioral economics book
		17	Mary Pat Campbell	Y	What I've Learned from the Good Judgement Project	summary of predictive modeling project under Office of Anticipating Surprise
		23	Brian Holland	Y	SVD of Weighted or Missing Data	how to use singular value decomposition when dealing with weighted or missing data
		25	Shea Parkes, Brad Armstrong	Y, Y	Calibrating Risk Score Model with Partial Credibility	description of ridge regression and its usage with specific portion of a larger data population
		30	Richard Xu, Dihui Lai, Minyu Cao, Scott Rushing, and Tim Rozar	Y	Appendix B: How to Build a Model	a single appendix from the research paper: Lapse Modeling for the Post-Level Period – a Practical Application of Predictive Modeling
		36	Doug Norris	Y	Simple Rating Systems: Entry-level sports forecasting	an intuitive approach to predictive modeling using progressive layers of sophistication
		40	Bryon Robidoux	Y	Stepping Out	an actuary's perspective on the 2015 Predictive Analyics World (PAW) conference
		43	Jeff Heaton	N[5]	What Big Data is, and How to Deal with It	tools and techniques (e.g., Vowpal Wabbit) for processing Big Data
		46	Charles Tsai	Y	A 'Hot Date' with Julia: Parallel Computations of Stochastic Valuations	description of a programming language with inherent parallel processing capability
		49	Dave Snell	Y	Index of all F&F articles: September, 2009 through July, 2015	this anthology

ENDNOTES

1 Electronic Only
2 Futurist
3 Underwriter
4 SOA Staff
5 Data Scientist
6 Behavioral Economist
7 Auditor
8 CSP
9 PhD Physics
10 Financial Planner
11 Student

Dave Snell

Dave Snell, ASA, MAAA, is technology evangelist at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *dave@ActuariesAndTechnology.com*

SOCIETY OF ACTUARIES Forecasting and Futurism Section

NONPROFIT
ORGANIZATION
U.S. POSTAGE
PAID
SAINT JOSEPH, MI
PERMIT NO. 263

Forecasting & Futurism

NEWSLETTER

475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
p: 847.706.3500 f: 847.706.3599
w: www.soa.org

MIX
Paper from responsible sources
FSC® C004755
www.fsc.org
FPO