Session 134: How to Get Real Results in Policyholder Behavior Modeling

SOA Antitrust Compliance Guidelines
SOA Presentation Disclaimer

How to Get Real Results in Policyholder Behavior Modeling

2019 ANNUAL MEETING & EXHIBIT

Rosmery Cruz and Timothy Paris

October 27, 2019

SOCIETY OF ACTUARIES®

SOCIETY OF ACTUARIES

Session Presented By:

Predictive Analytics and Futurism Section

Provides opportunities for actuaries to deepen their understanding of predictive analytics and emerging artificial intelligence (AI) and data science methods relevant to the insurance industry.

Section Developed Content & Benefits

The Predictive Analytics and Futurism Newsletter

Published three times a year. Members receive a hard copy by mail and have access to a digital edition.

Research

Research on topics, including the validation of predictive models and a Delphi study on economic scenario generation.

SOA Meetings and Seminars

Section developed content presented during meeting sessions and seminars.

Webcasts & Podcasts

Discounts on section developed webcasts. Free access to podcasts and section created webcasts over one-year old.

Join the PAF Section Today! ***SOA.org/PAF***

SOCIETY OF ACTUARIES
Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- Do not discuss prices for services or products or anything else that might affect prices
- Do not discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- Do not speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- Do leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- Do alert SOA staff and/or legal counsel to any concerning discussions
- Do consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

Agenda

01 Motivation

02 Overfitting: What & Why

03 Case Study: Variable Annuity Surrender Rates

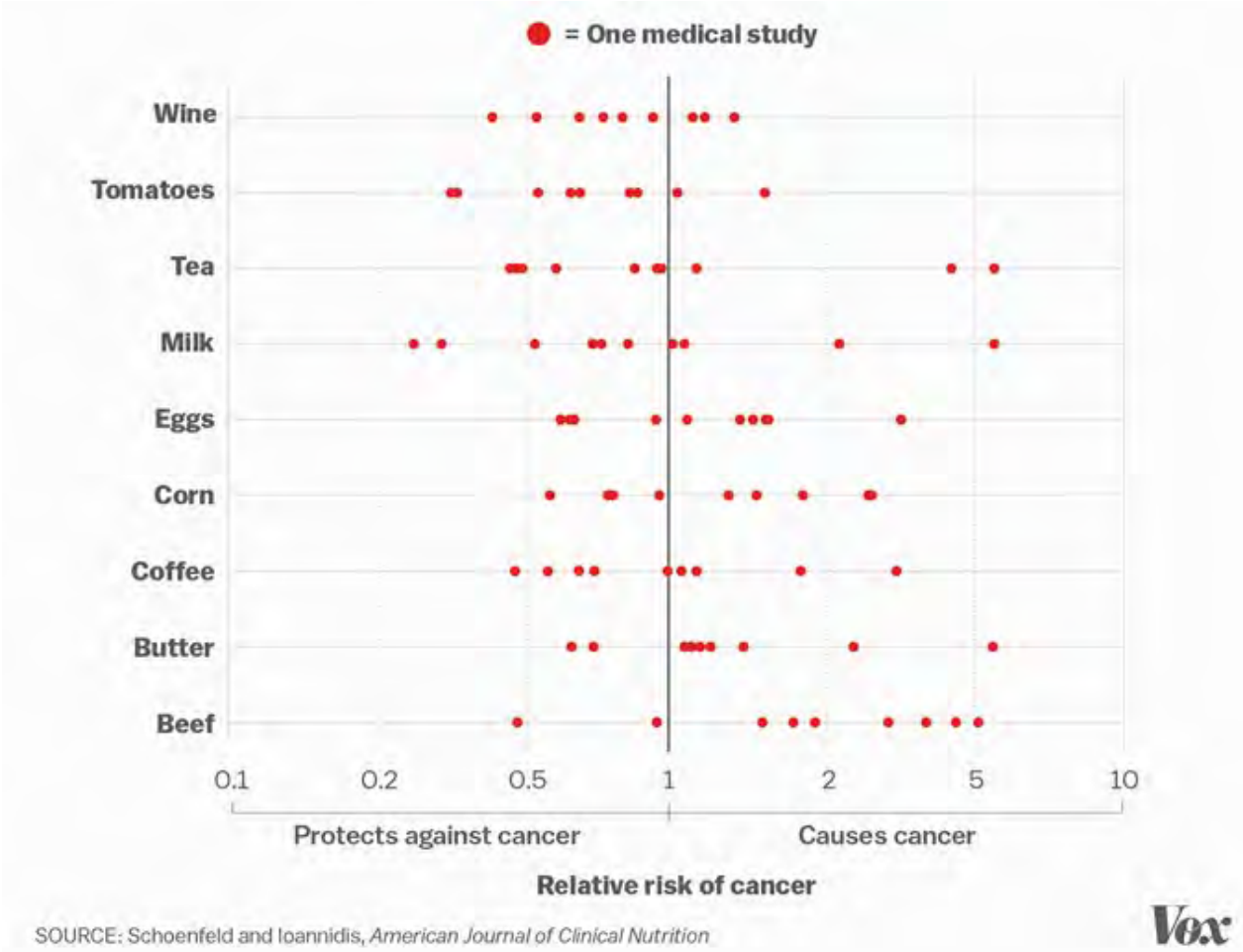
04 Learnings

Motivation

Published studies featured in the media

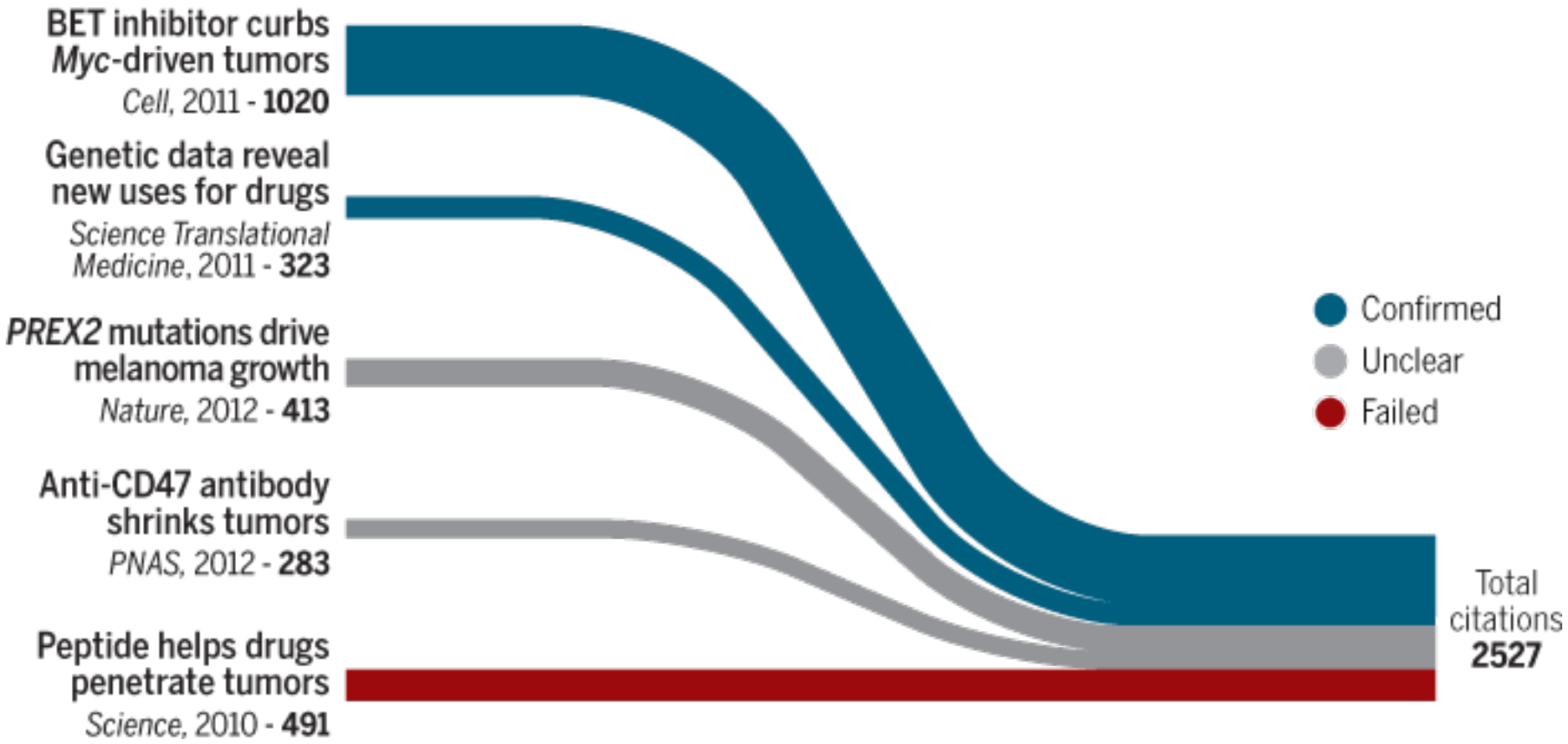
- “Late-night eating hurts learning and memory”
- “Science proves pizza is the most addictive food”
- “A glass of red wine a day can equal to an hour in the gym”
- “Driving dehydrated just as dangerous as driving drunk”

Everything we eat both causes and prevents cancer



The American Journal of Clinical Nutrition, Volume 97, Issue 1, January 2013, Pages 127–134, https://doi.org/10.3945/ajcn.112.047142

Rigorous replication effort succeeds for just two of five cancer papers



Science, "Rigorous replication effort succeeds for just two of five cancer papers," http://www.sciencemag.org/news/2017/01/rigorous-replication-effort-succeeds-just-two-five-cancer-papers accessed August 18, 2018.

Single medical studies by the numbers

6%

Of new journal articles reviewed annually are deemed high-quality enough to inform patient care

SOURCE: Haynes, Evidence Based Nursing

29%

Of highly cited original medical studies were either contradicted by later studies or were found to have much smaller effects than original articles suggested

SOURCE: Ioannidis, JAMA

5

Only 5 Of 101 new therapies or medicines claimed by medical studies to be promising made it to market

SOURCE: Contopoulos-Ioannidis, American Journal of Medicine

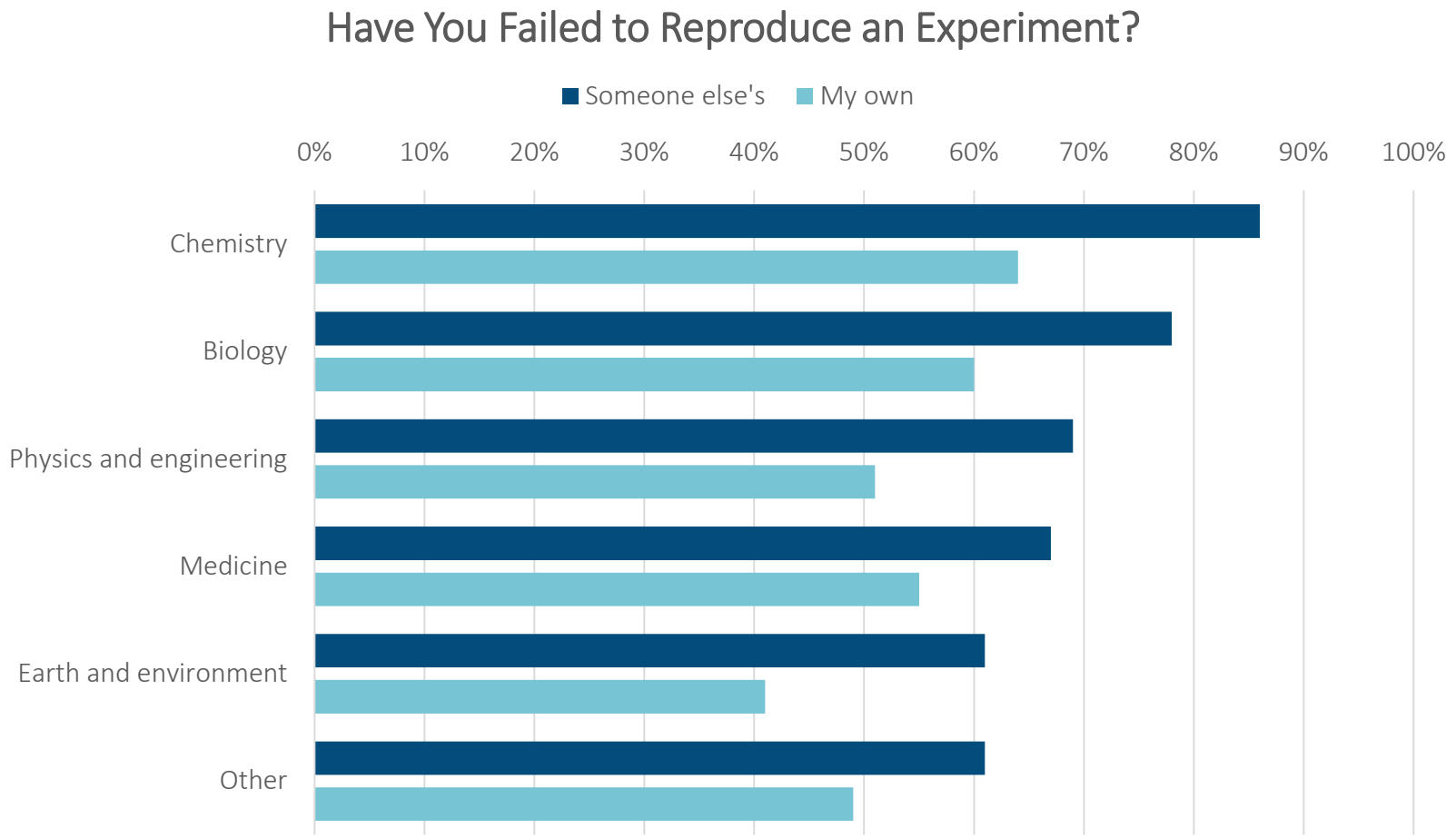
$200B

Of annual global spending on research is wasted on badly designed or redundant studies

SOURCE: Macleod, Lancet

Belluz, J. (2017, February 27). *This is why you shouldn't believe that exciting new medical study*. https://www.vox.com/2015/3/23/8264355/research-study-hype

Most scientists have experienced failure to reproduce results



Baker, M. (2016, May 25). *1,500 scientists lift the lid on reproducibility*. https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

Publication Asymmetry

- Once something appears in print, it becomes very difficult to criticize
- Incentives to publish positive replications are low
- Journals can be reluctant to publish negative findings

Dietvorst, B., Simmons, J. P., & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. Journal of Experimental Psychology: General, 144 (1), 114-126. http://dx.doi.org/10.1037/xge0000033

Major medical journals don’t follow their own rules for reporting results from clinical trials

- Editors and researchers routinely misunderstand what correct trial reporting looks like
- Authors should describe the outcomes they plan to study before a trial starts and stick to that list when they publish the trial
- This varied by journal

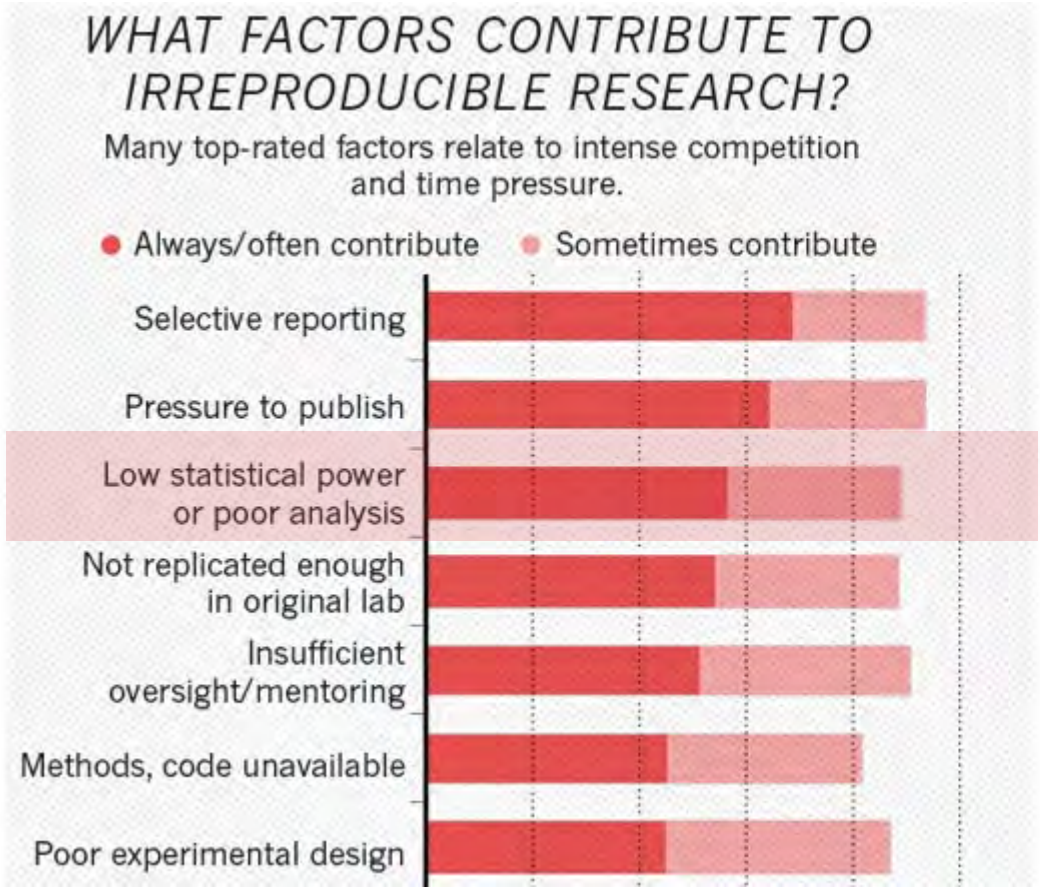
9 out of 67 Trials published in the five journals reported outcomes correctly, the COMPare team reported on 14 February in the journal *Trials*.

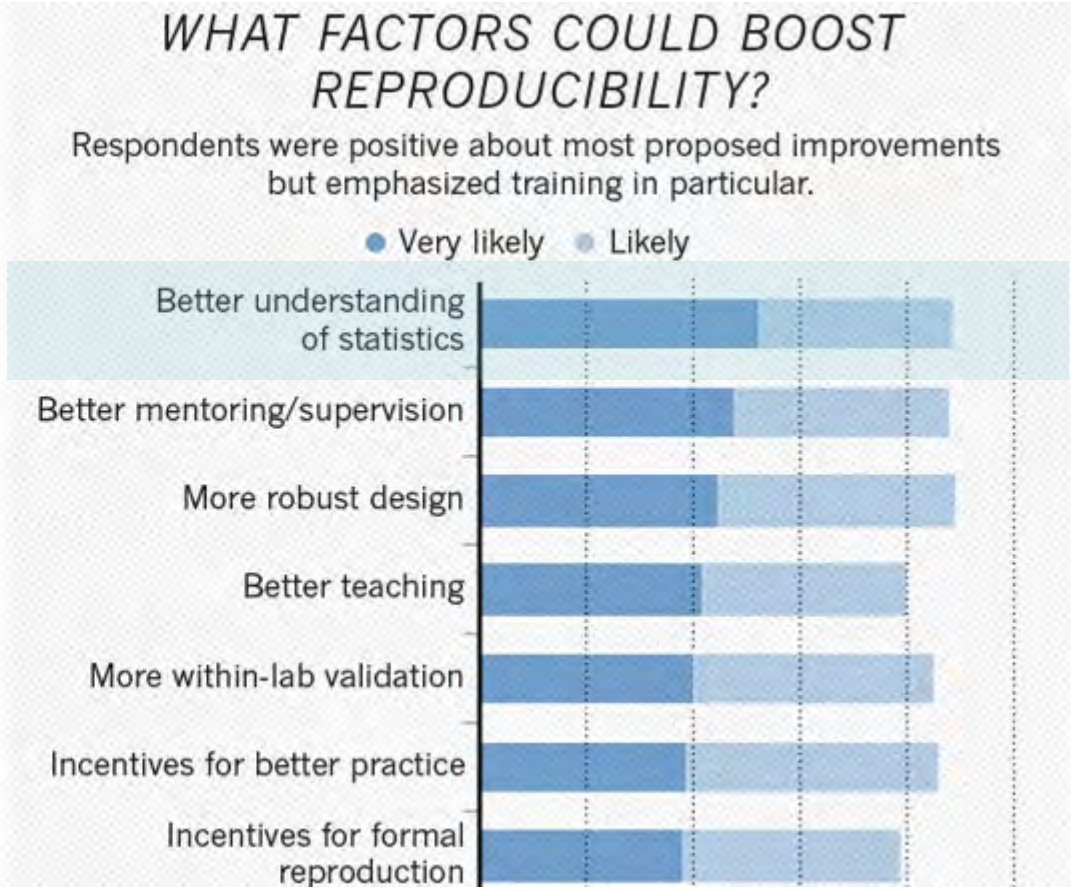
25% Didn’t correctly report the primary outcome they set out to measure and

45% Didn’t properly report all secondary outcomes

Kaiser, J. (2019, February 15). *Major medical journals don’t follow their own rules for reporting results from clinical trials*. https://www.sciencemag.org/news/2019/02/major-medical-journals-don-t-follow-their-own-rules-reporting-results-clinical-trials

Reasons for the Replication Crisis



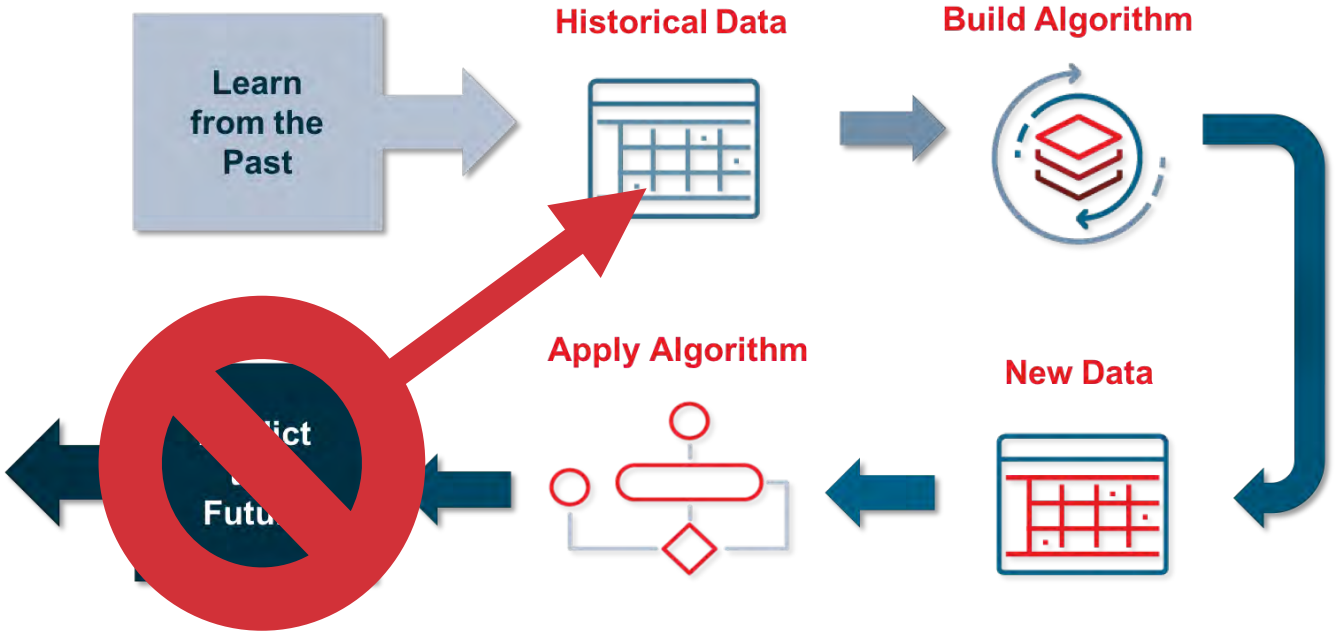


Baker, M. (2016, May 25). *1,500 scientists lift the lid on reproducibility*. https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

Overfitting: What & Why

Overfitting Definition

"The problem of capitalizing on the idiosyncratic characteristics of the sample at hand. Overfitting yields overly optimistic model results: "findings" that appear in an overfitted model don't really exist in the population and hence will not replicate." (Babyak, 2004)

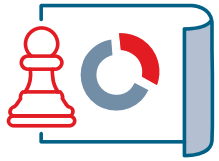


One of many definitions

Text from *Babyak 2004: What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models.*

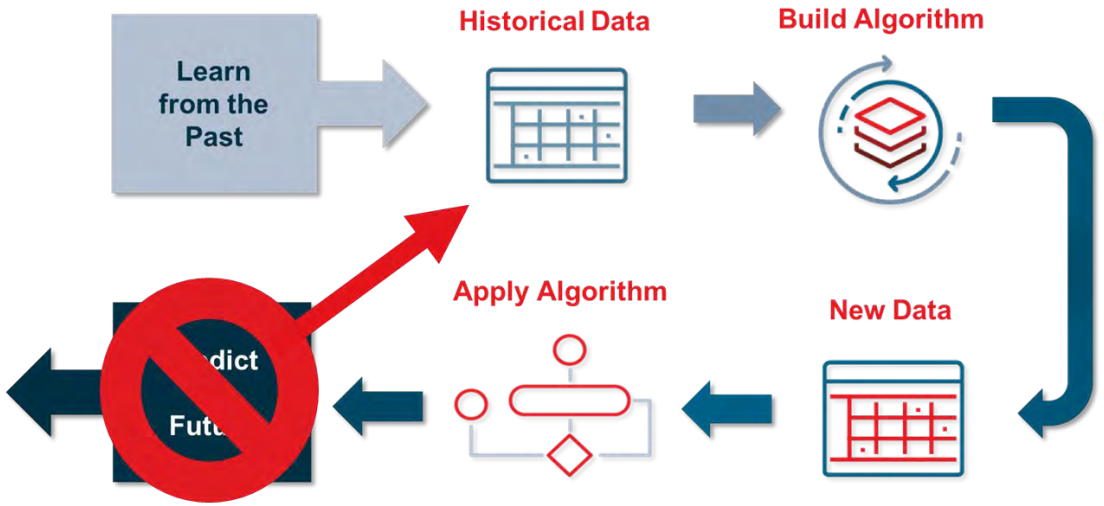
When Does Overfitting Occur?

Generally, overfitting occurs due to analyst oversight in two key areas:



Researcher degrees of freedom (also known as procedural overfitting, data dredging, p-hacking, etc.)

Asking too much from the data (model complexity)



The Garden of Forking Paths

Forking paths come from choices in data processing and also from choices in analysis

- A group of researchers plans to compare three dosages of a drug in a clinical trial.
- There's no pre-planned intent to compare effects broken down by sex, but the sex of the subjects is routinely recorded.
- They have informally made fifteen comparisons

Dietvorst, B., Simmons, J. P., & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. Journal of Experimental Psychology: General, 144 (1), 114-126. http://dx.doi.org/10.1037/xge0000033

The Garden of Forking Paths

BRITAIN'S MOST TRUSTED NATIONAL NEWSPAPER

THE TIMES

Friday January 4 2019 | thetimes.co.uk | No 72735

£1.60 Subscribers save 50p

Bricks & Mortar

House prices in 2019

What's going up, what's coming down Pullout

Plus Another fine mess? The new Laurel and Hardy Times2

£200,000 spent on protecting hate preacher's human rights

Paul Morgan-Bentley
Head of Investigations

Britain has spent almost £200,000 protecting the welfare of the hate preacher Abu Qatada since he was deported to Jordan in 2013, The Times can reveal.

Under terms agreed by Theresa May when she was home secretary, the government has paid for the cleric to have appointments with human rights workers and doctors for three years. The payments were agreed despite Mrs May telling parliament in 2013 that "significant costs" to the taxpayer relating to the Abu Qatada case were "not acceptable to the public and not acceptable to me".

The "welfare visits" were to ensure that he was not tortured after he was removed from Britain for being a threat to national security, details released under freedom of information laws show.

The sums spent to secure Abu Qatada's removal prompted an outcry. They included at least £1.7 million in legal fees, £640,000 of which covered his legal aid. The total cost of his deportation, including the welfare payments, is thought to have exceeded £1.9 million. MPs and campaigners said yesterday it was "ridiculous" and "an insult to taxpayers" that he had continued to benefit from welfare payments.

The fees are a legacy of the policy of deportation with assurances (DWA) that was favoured by Mrs May, which allows the removal of suspected terrorists to their home countries with guarantees that their rights will not be infringed. A 2017 report criticised it for being too expensive and failing to effect the removal of any suspects since 2013.

Abu Qatada, whose real name is Omar Mahmoud Mohammed Othman, was given asylum in Britain in 1994 before being detained as a terror suspect after the September 11, 2001 attacks in the United States. He was

Continued on page 2, col 3

Doctors tell parents to cut children's screen time

Concern over social media link to depression

Chris Smyth Health Editor

Doctors have issued the first guidance advising parents to limit their children's access to technology as a study linked heavy social media use by teenagers to signs of depression.

Children should not watch television or go online within an hour of bedtime, doctors have recommended. Parents should also set a good example by controlling their own phone use in front of children, the Royal College of Paediatrics and Child Health said.

The guidance came as a study found that teenagers who spent long hours on social media were twice as likely to suffer mental health than had been seen with other screen time, such as watching television. "I suspect social media is a case apart from other screens because of its interactive nature."

The advice from the college is that although screens are not inherently bad, long hours online or watching televisions risk distracting children from sleep, exercise and family time.

Blue light from screens is thought to interfere with production of the sleep-inducing hormone melatonin, while overstimulation also keeps children awake. Poor sleep is known to increase the risk of depression.

A quarter of girls in their late teens

British citizen held in Russia over 'spying for the West'

Catherine Philp
Diplomatic Correspondent
Tom Parfitt Moscow

A British citizen has been detained in Russia on suspicion of spying, The Times has learnt.

Paul Whelan, 48, was formally charged with espionage yesterday as it was alleged that he had received in a Moscow hotel room a computer memory stick containing a secret list of Russian agents.

Mr Whelan, a former US Marine, is also a US citizen. American embassy officials in Moscow contacted their British counterparts to inform them of Mr Whelan's arrest and his status as a British citizen.

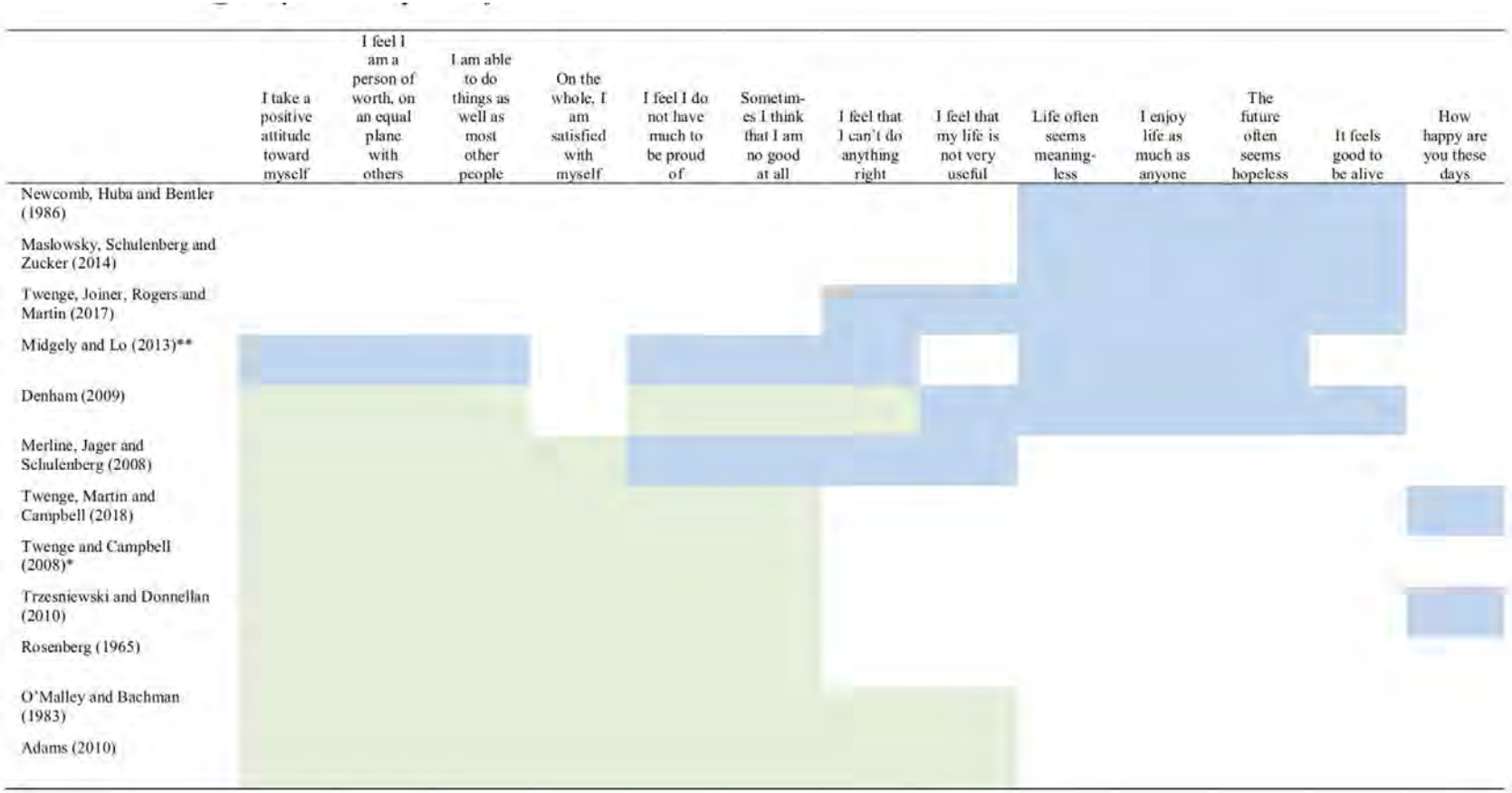
A Foreign Office spokesman said: "Staff have requested consular access to a British man detained in Russia after receiving a request for assistance."

The disclosure will add tension to relations between Moscow and London, already fraught over the poisoning of the Russian military intelligence officer Sergei Skripal, and his daughter, Yulia, in Salisbury last March.

Mr Whelan was seized at the Metropol hotel minutes after the storage device was passed to him, an unidentified

Several studies published on the association between adolescent well-being and digital reported by many news outlets

Scientists could have analyzed the data in over a trillion ways

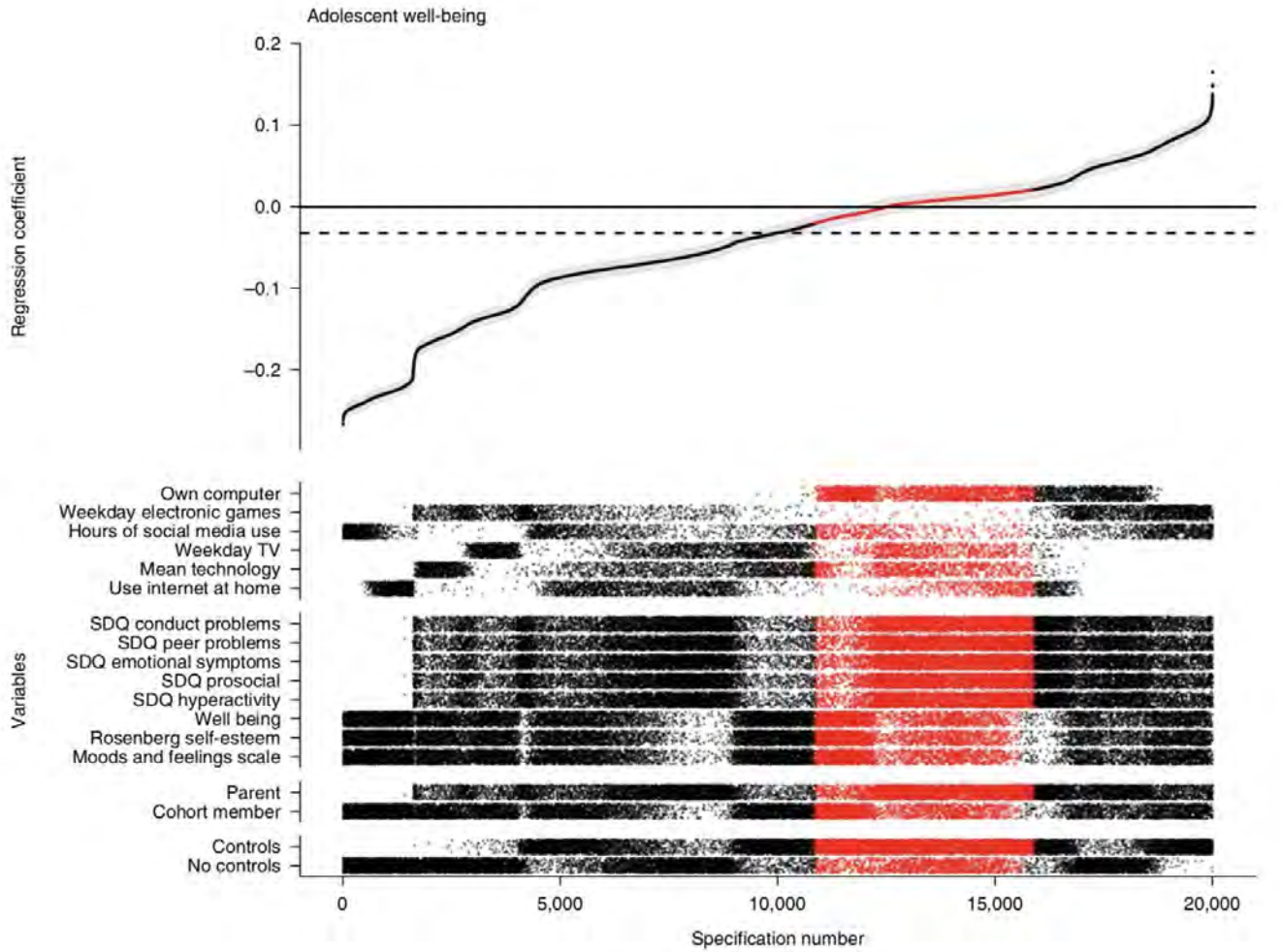


Differences in:

- How to define well-being
- How to define technology use
- Model specifications
- …etc.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. Nature Human Behaviour, 3, 173-182.

Number of (Plausible) Forking Paths: 603,979,752

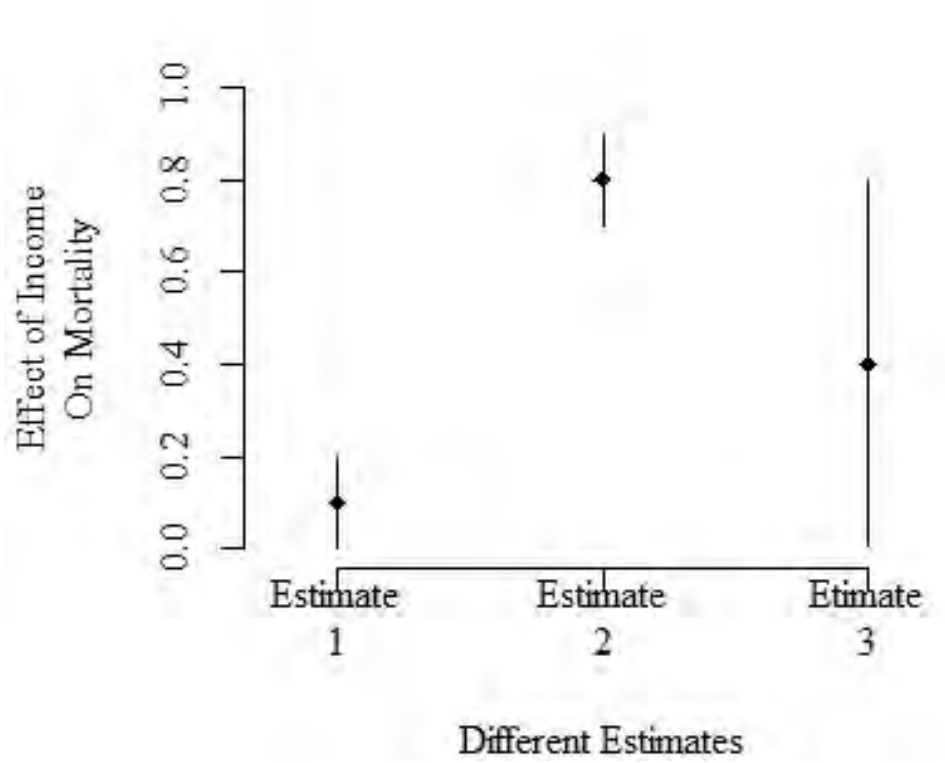


"The association we find between digital technology use and adolescent well-being is negative but small, explaining at most 0.4% of the variation in well-being."

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. Nature Human Behaviour, 3, 173-182.

The Problem With Statistical Significance

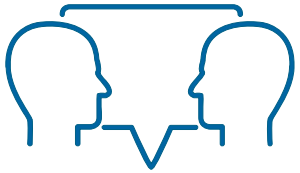
- “Significantitis” or “Dichotomania” (Greenland, 2017)
- Overreliance on phrases like “We deemed a p value less than 0.05 to be significant,”
- P-values are extremely noisy unless underlying effect is huge



Greenland, S. (2017). The need for cognitive science in methodology. American Journal of Epidemiology 186, 639–645

When Does Overfitting Occur?

Make research design decisions before analyzing the data



Where applicable, use subject matter knowledge to inform data aggregation (i.e., age groups)

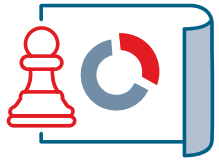
Limit the exclusion of data

Validate your results (discussed later in the presentation)

Strategies to Minimize Researcher Degrees of Freedom

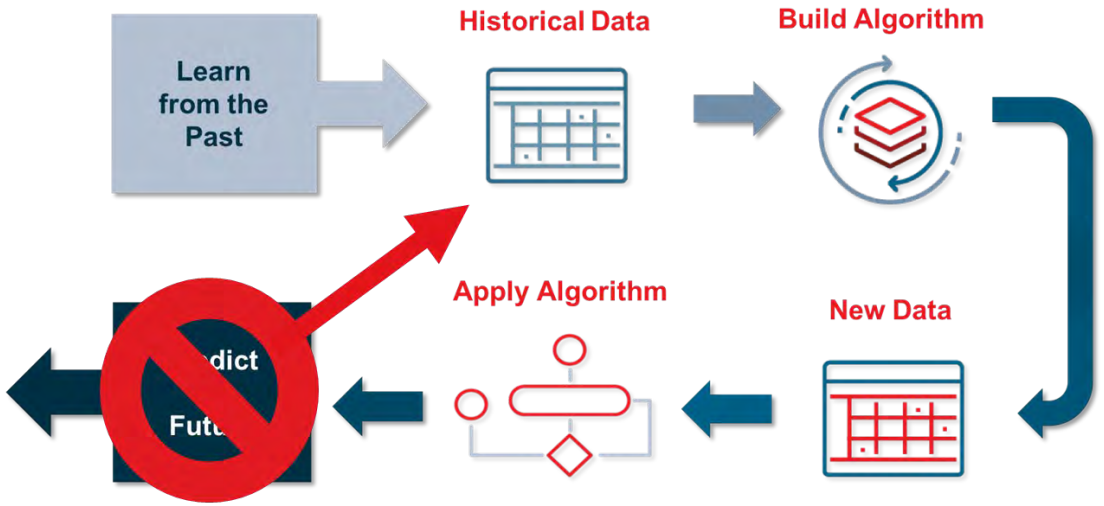
When Does Overfitting Occur?

Generally, overfitting occurs due to analyst oversight in two key areas:



Researcher degrees of freedom (also known as procedural overfitting, data dredging, p-hacking, etc.)

Asking too much from the data (model complexity)



When Does Overfitting Occur?

"Given a certain number of observations in a data set, there is an upper limit to the complexity of the model that can be derived with any acceptable degree of uncertainty." (Babyak, 2004)

Asking too much of the data

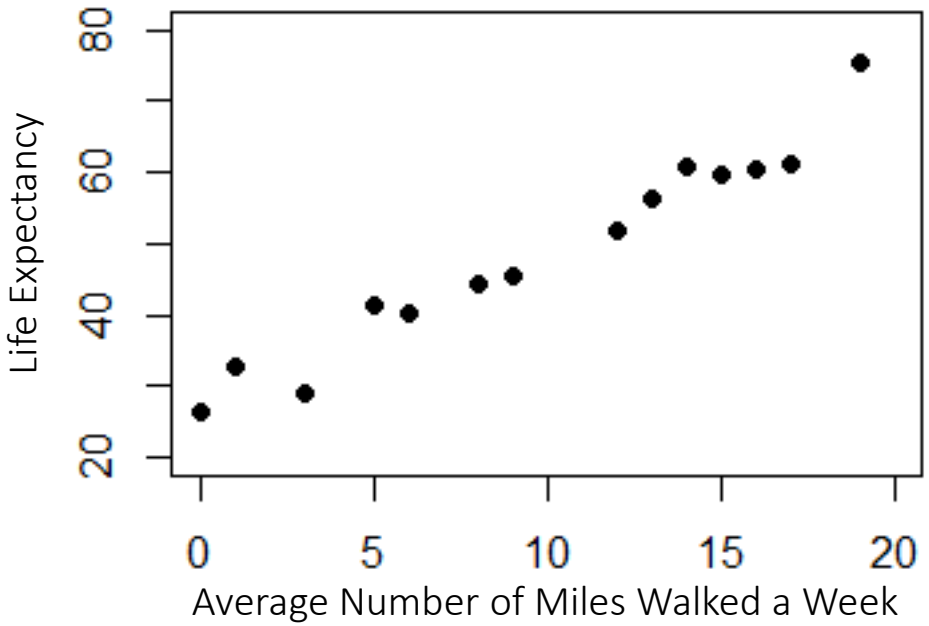
Text from Babyak 2004: What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models.

How Do You Prevent Overfitting?

1

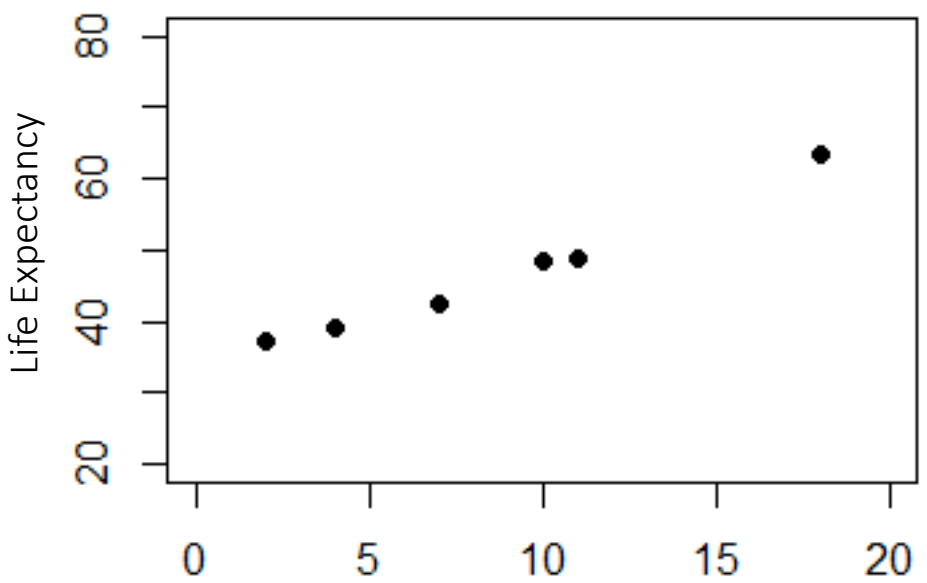
Training Data

N = 14



Test Data

N = 6

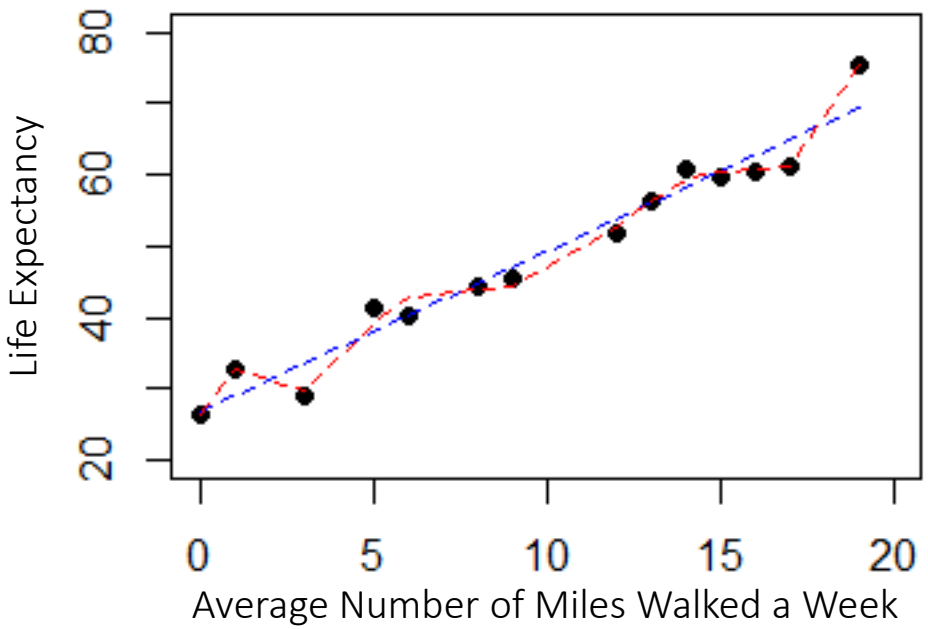


How Do You Prevent Overfitting?

1

Training Data

N = 14

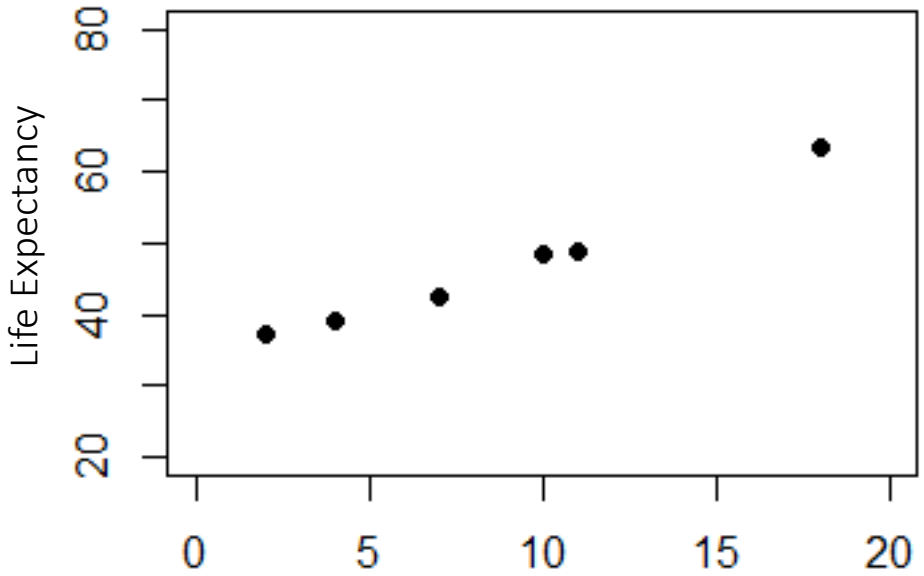


Simple Model Training MSE: 8.21

Complex Model Training MSE:1.24

Test Data

N = 6



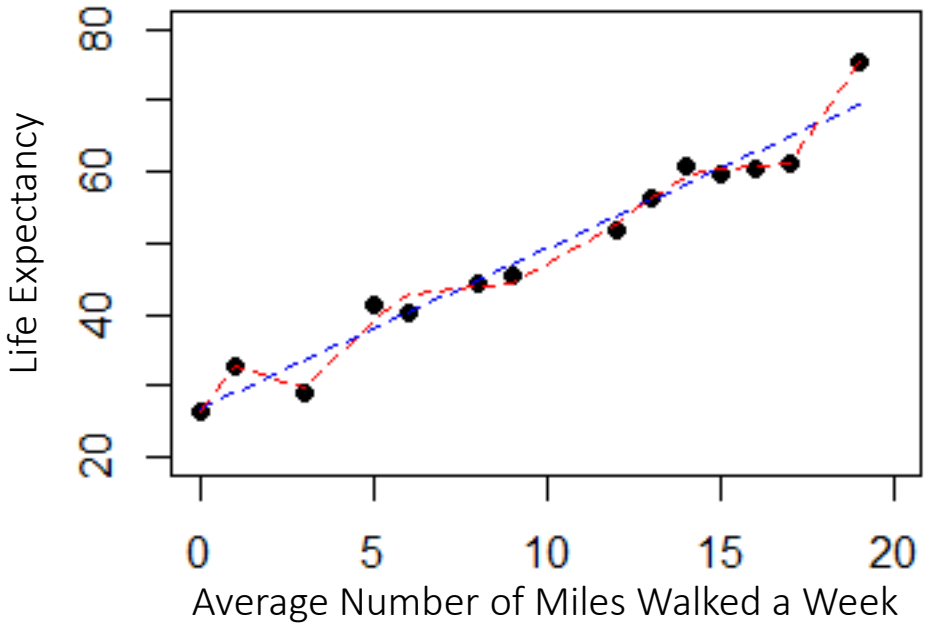
2

How Do You Prevent Overfitting?

1

Training Data

N = 14



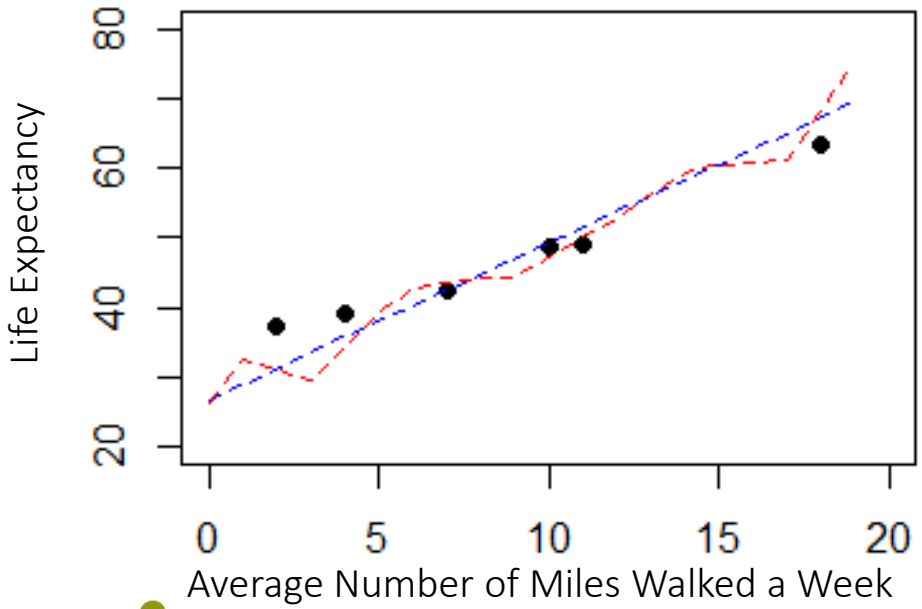
Simple Model Training MSE: 8.21

Complex Model Training MSE:1.24

2

Test Data

N = 6



Simple Model Test MSE: 11.60

Complex Model Test MSE: 16.95

3

How Do You Prevent Overfitting?

01 Test-set

02 Cross-Validation

03 Leave-one-out Cross Validation

These are some additional classical ways to approach overfitting and researcher degrees of freedom:

- AIC/BIC metrics
- Bootstrapping
- Bonferroni correction (adjusts for multiple comparisons)

Case Study:
Variable Annuity Surrender Rates

SOCIETY OF ACTUARIES®

VM-21 PBR for Variable Annuities

Public redline exposure draft as of April 30, 2019
https://naic-cms.org/exposure-drafts

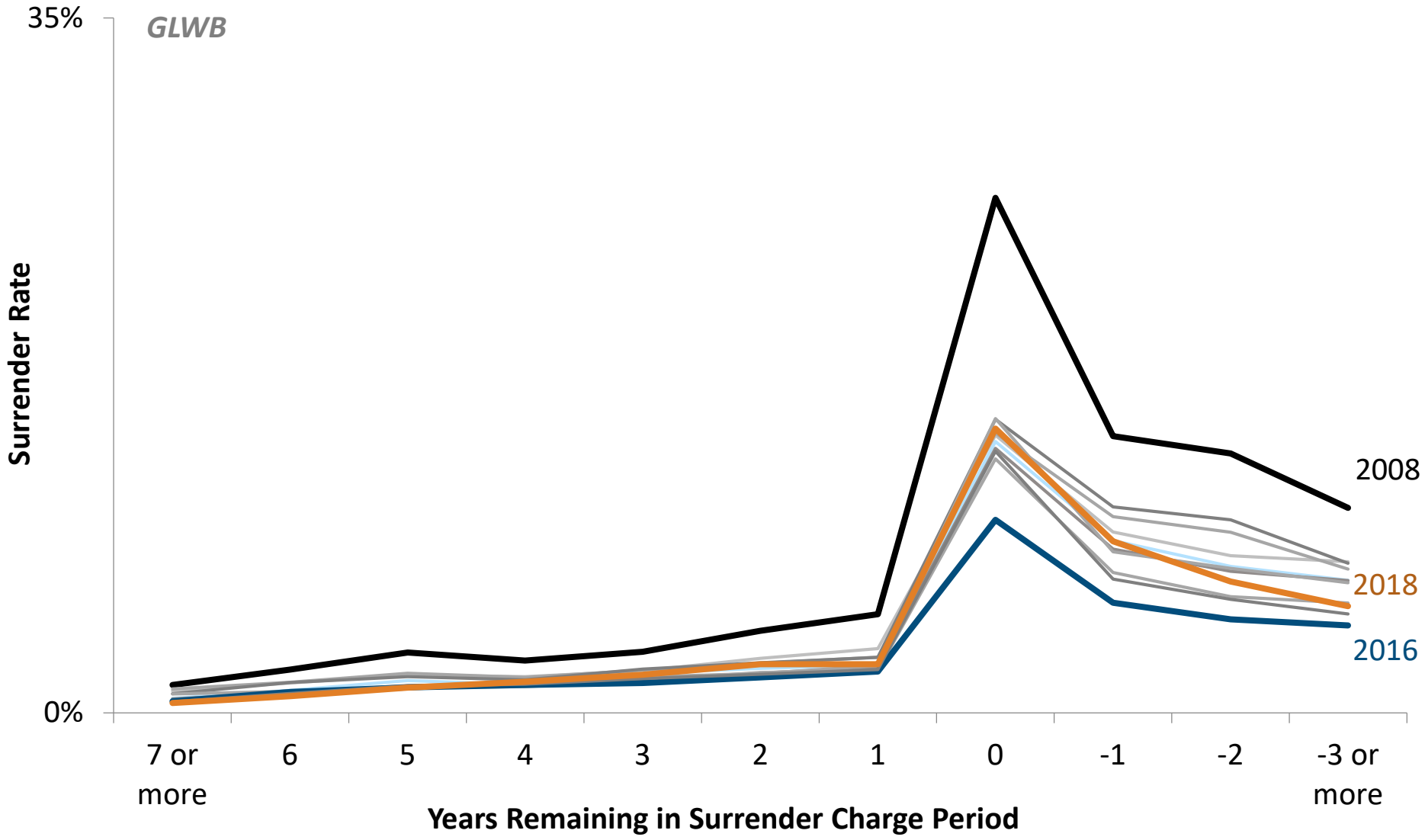
Section 10: Contract Holder Behavior Assumptions

1. Should examine many factors including cohorts, product features, distribution channels, option values, rationality, static vs dynamic
2. Required sensitivity testing, with margins inversely related to data credibility
3. Unless there is clear evidence to the contrary, should be no less conservative than past experience and efficiency should increase over time
4. Where direct data is lacking, should look to similar data from other sources/companies

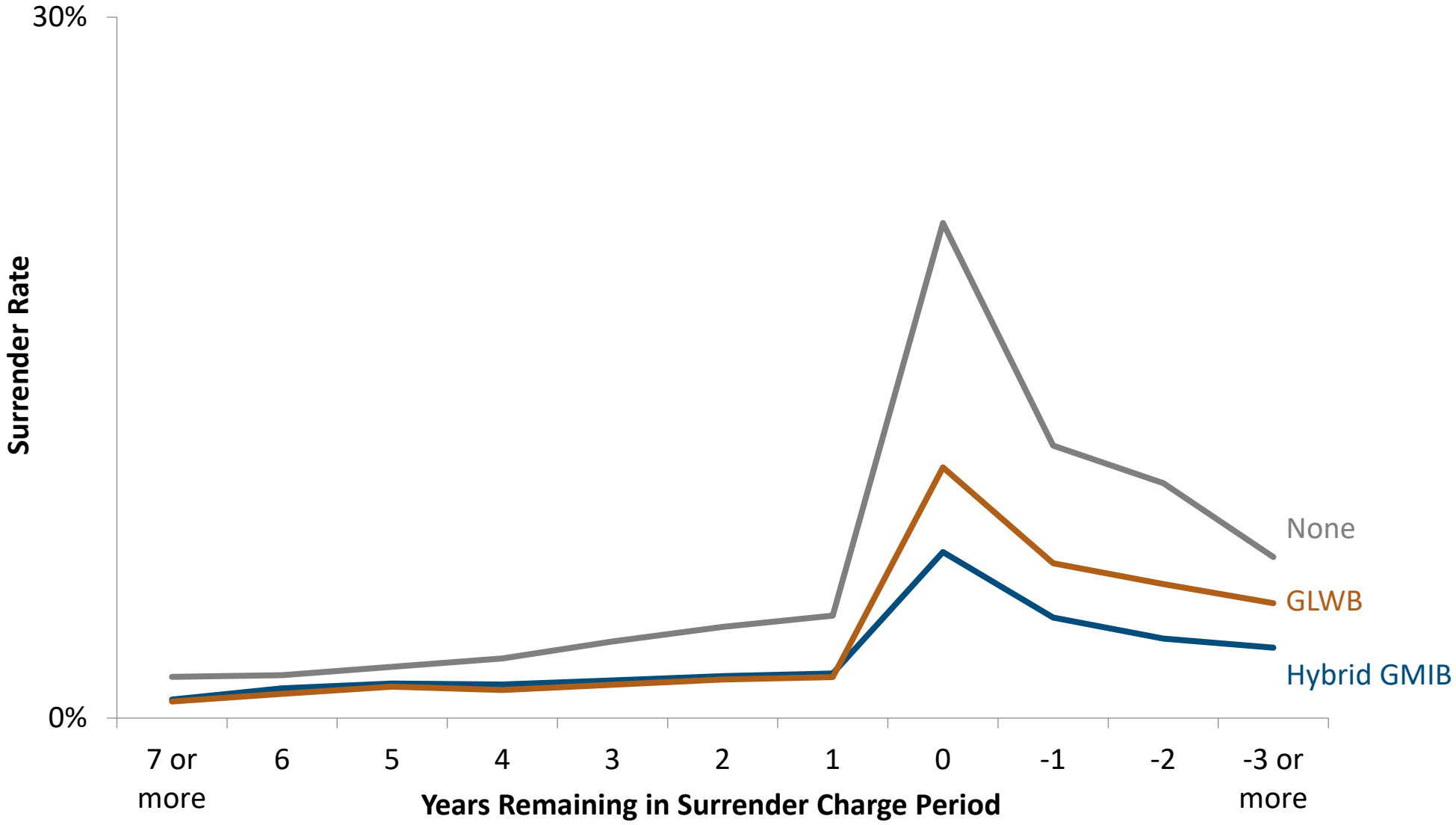

You and your data
Building models with your data
Improving models with industry data

You and Your Data

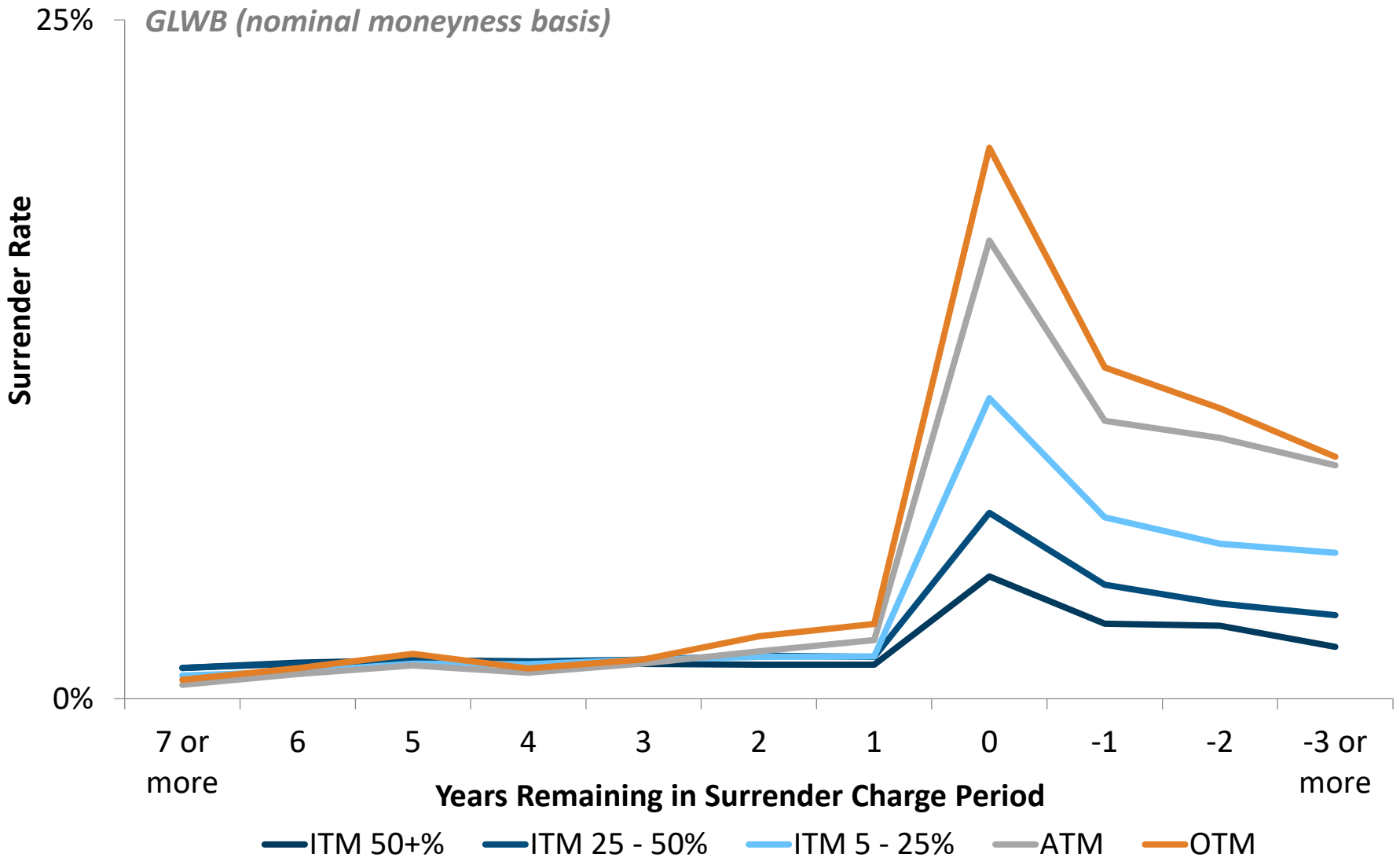
Surrender charges work, but impact has changed over the years



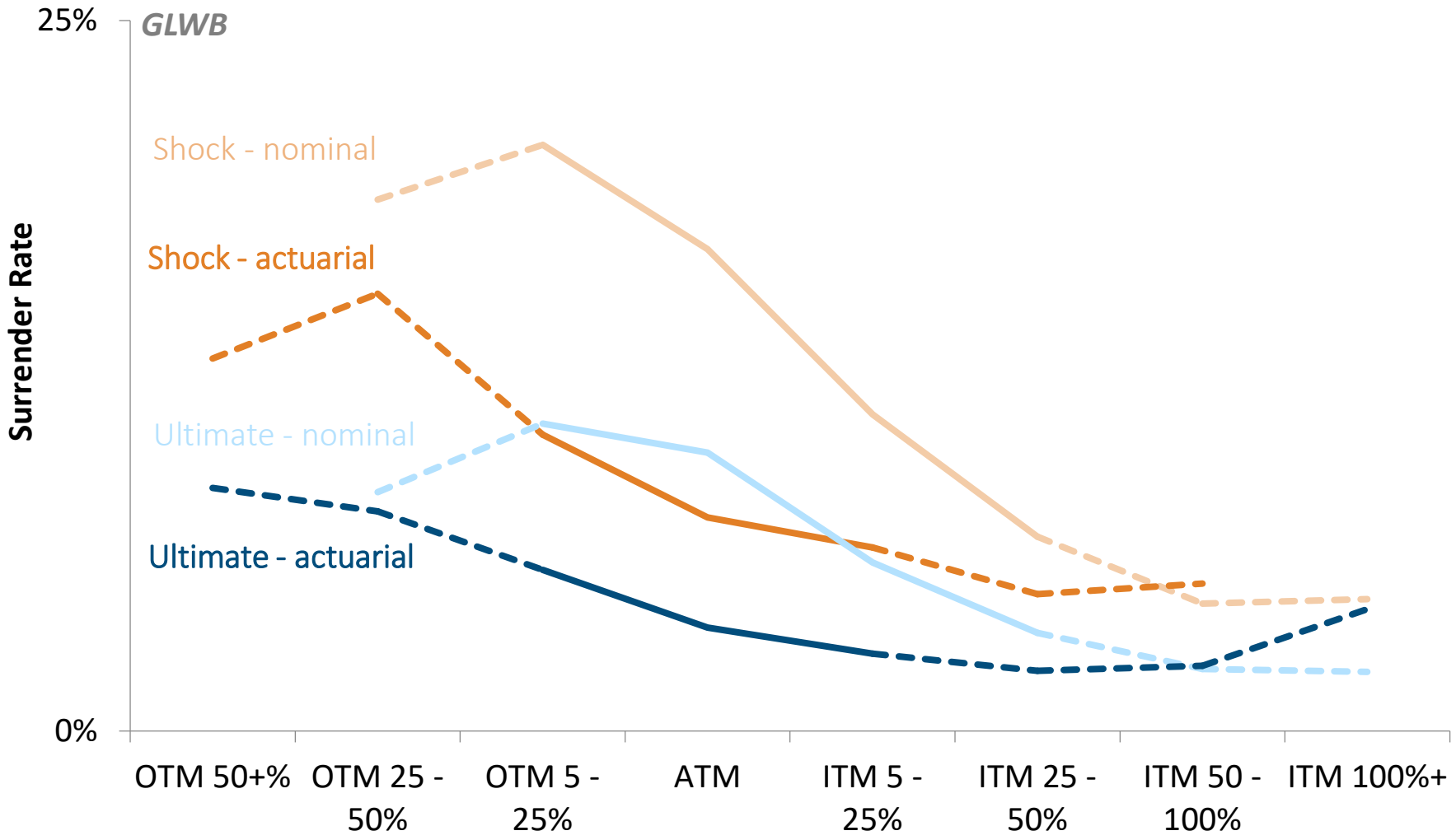
Surrender rates are lower with living benefit guarantees…



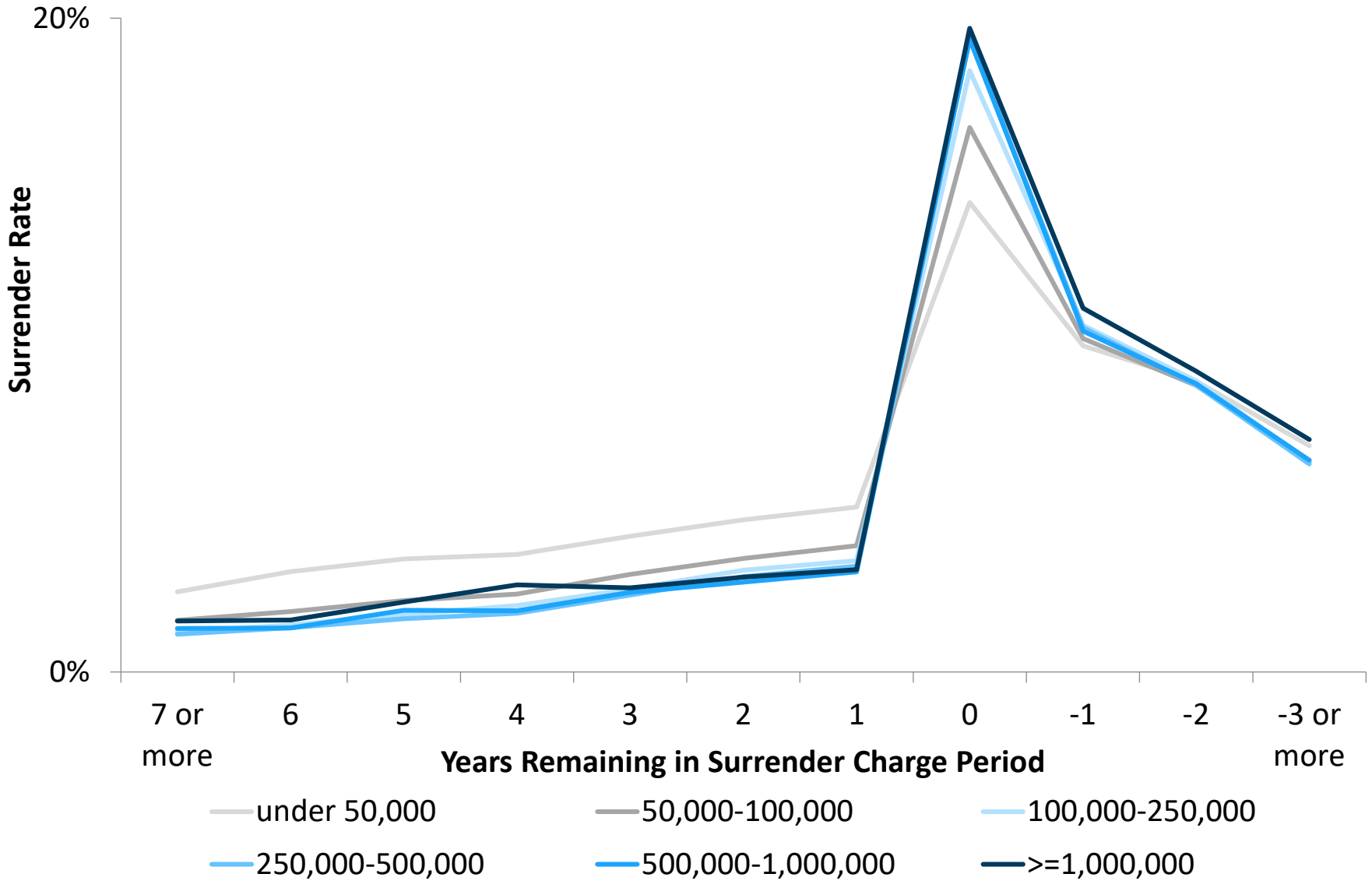
…and when guarantees are more valuable



How you measure value matters, but company-level credibility is very limited



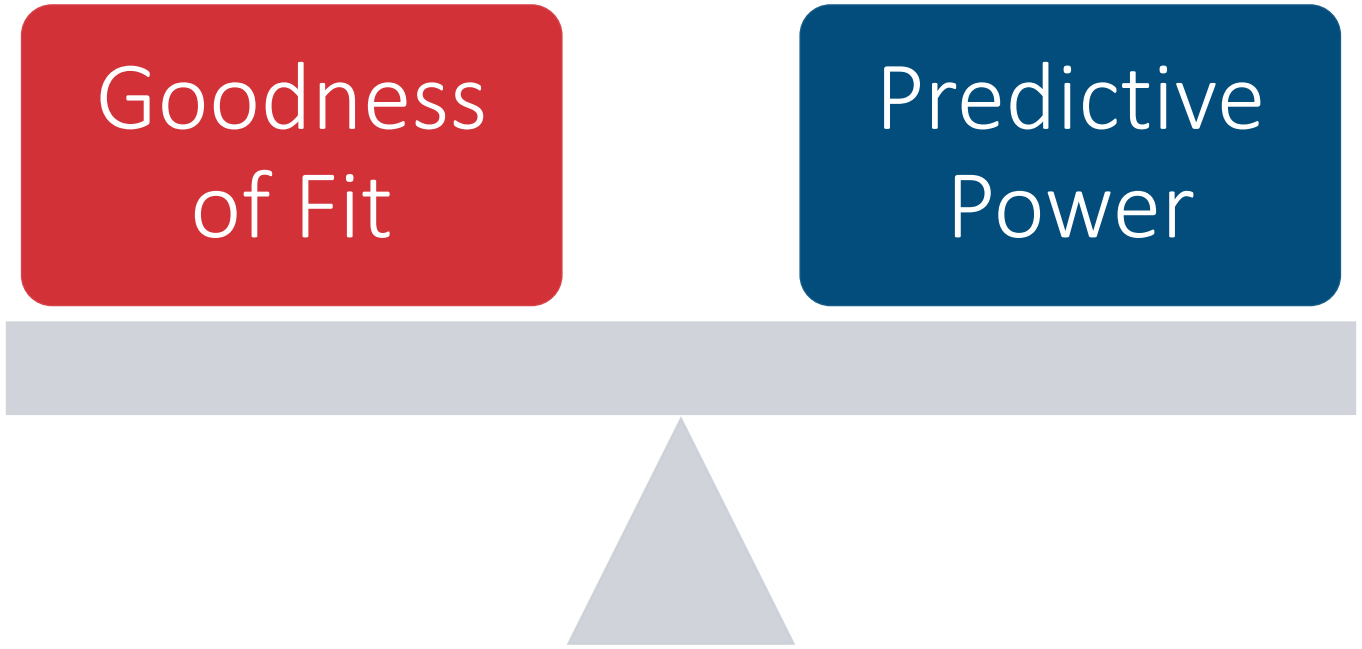
Largest and smallest contracts behave differently

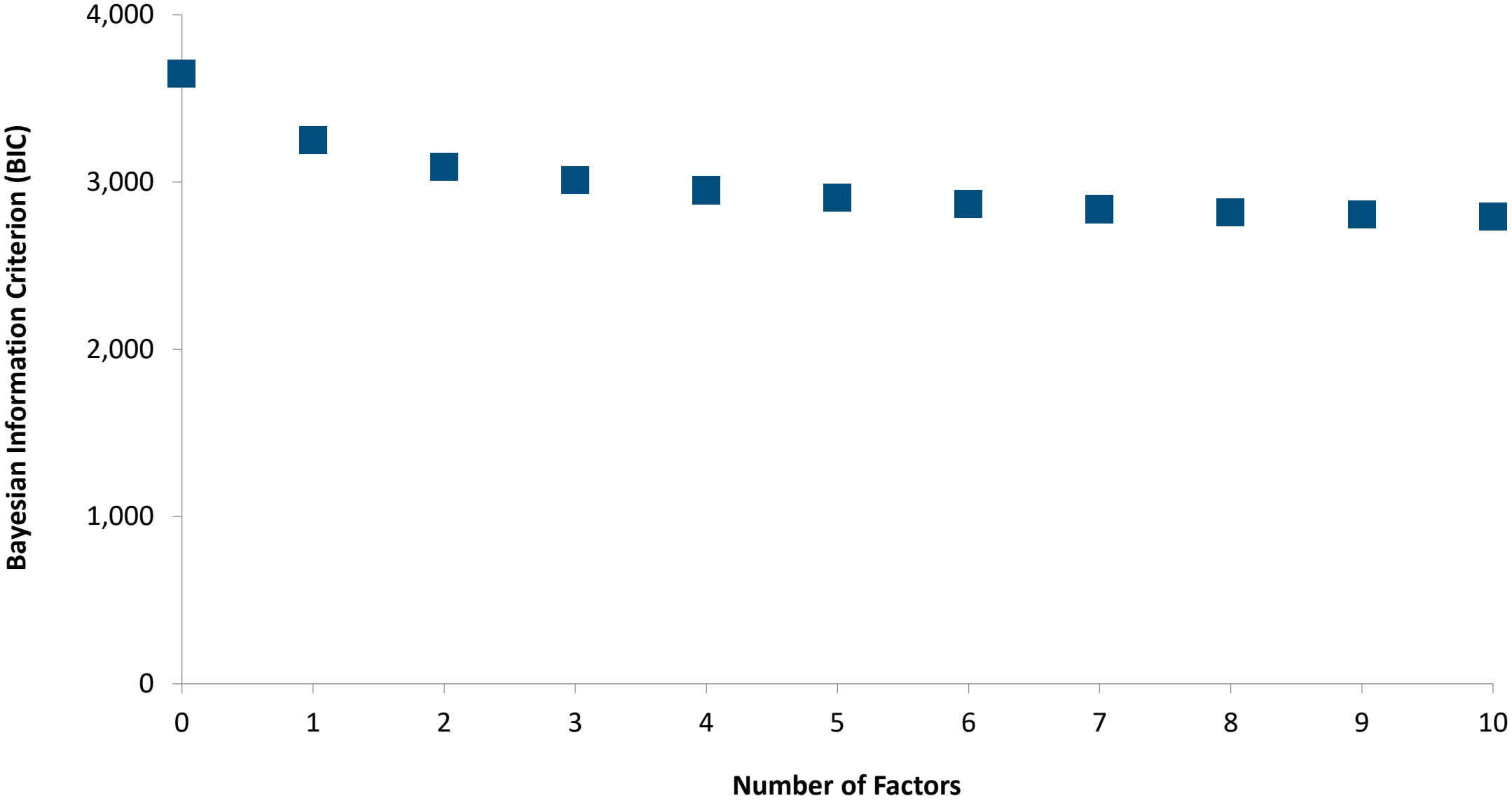


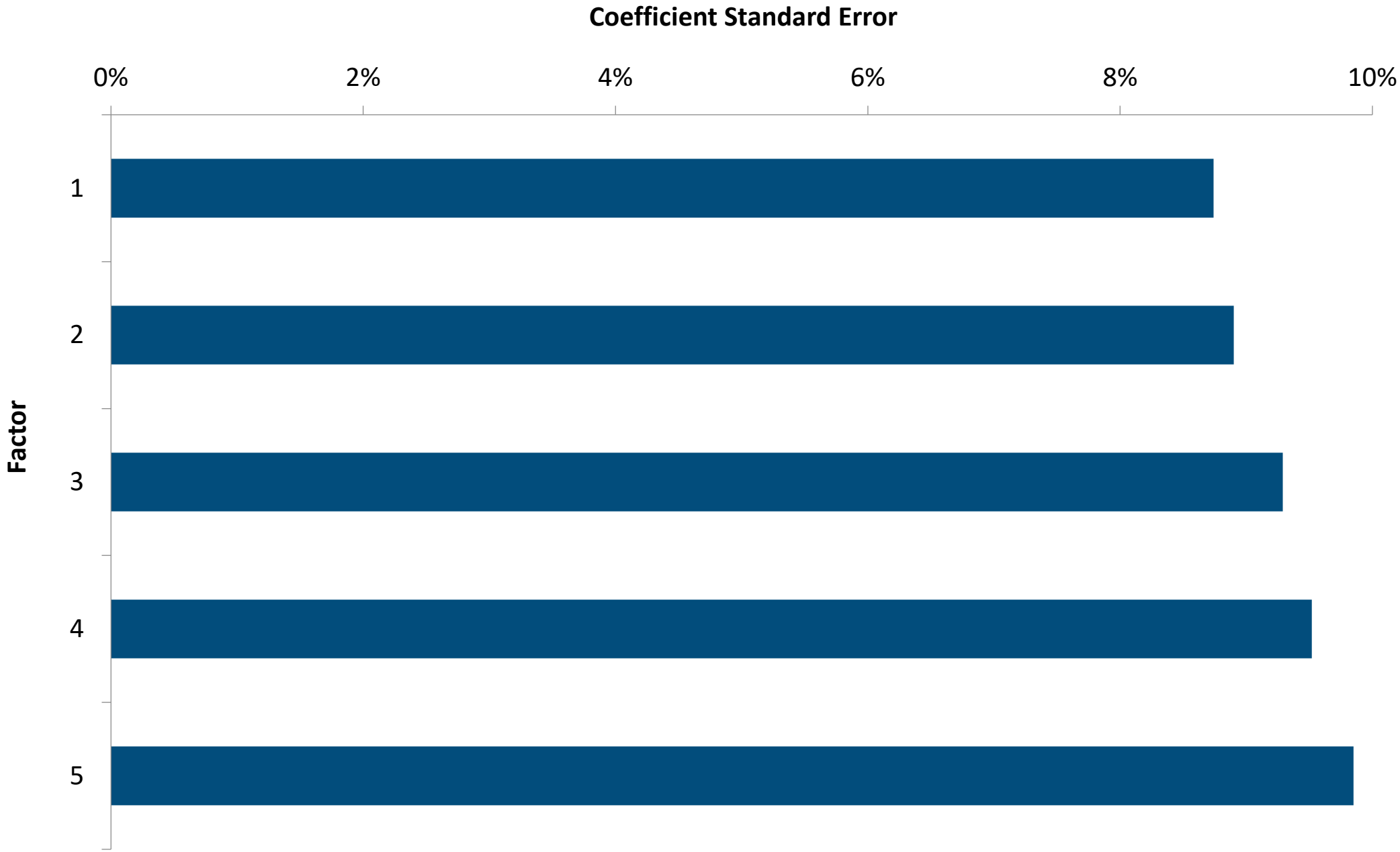
Building Models with Your Data

Modeling and assumptions

- Measuring goodness-of-fit for candidate models
- Testing predictive power on out-of-sample data
- Art + science: choosing, communicating, and ongoing recalibration

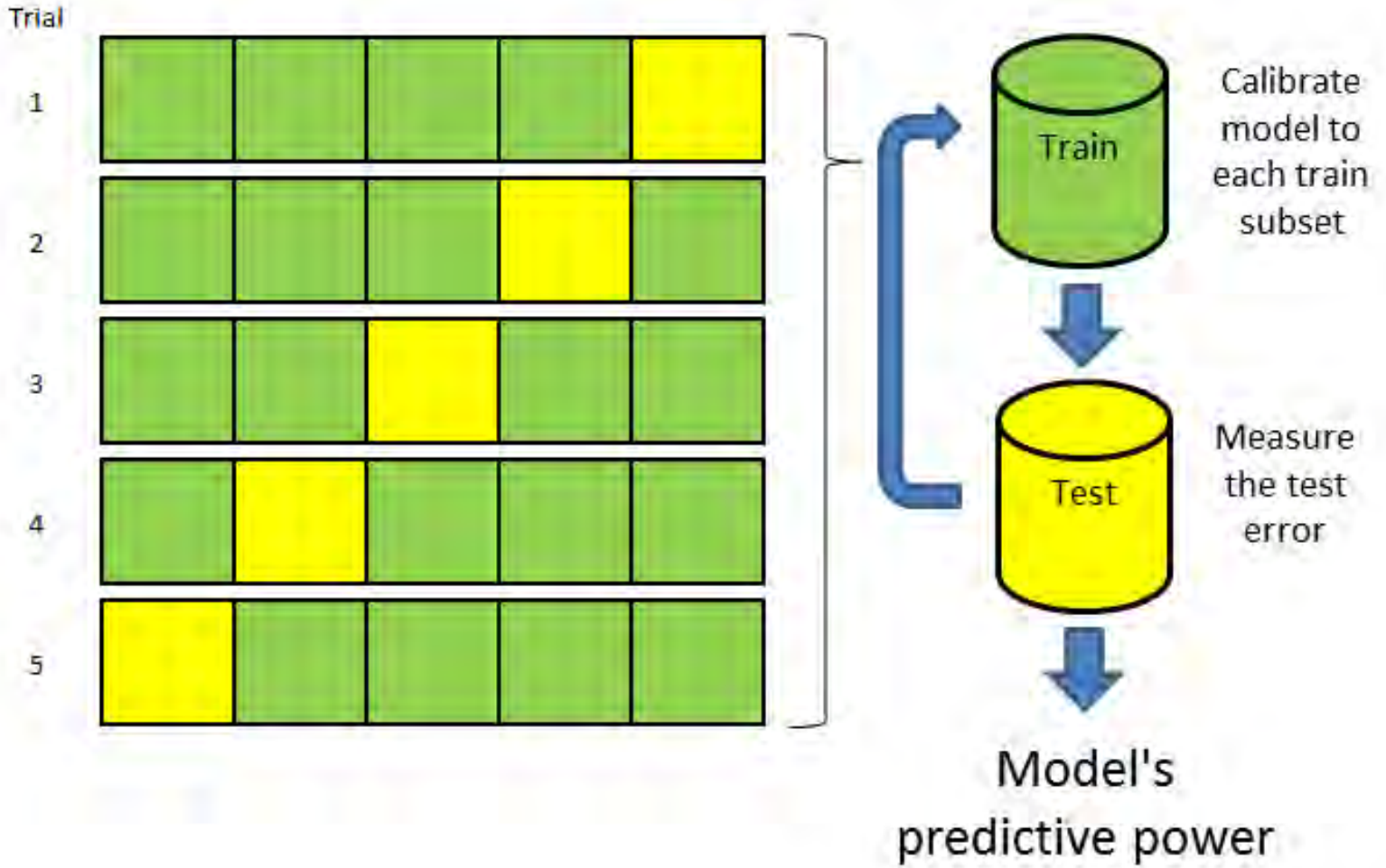

Goodness of Fit
Predictive Power

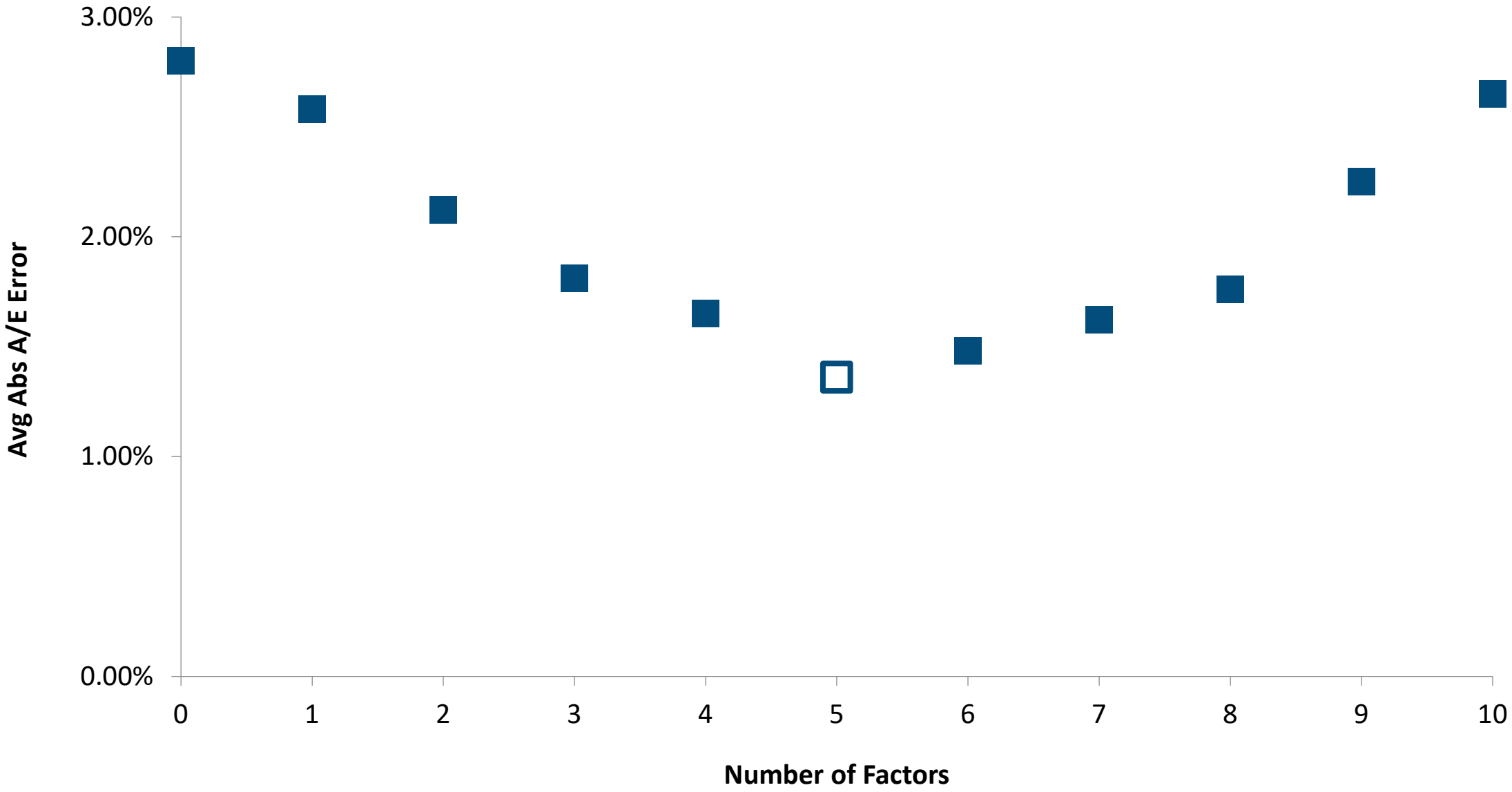

4,000
3,000
2,000
1,000
0
Bayesian Information Criterion (BIC)
0
1
2
3
4
5
6
7
8
9
10
Number of Factors

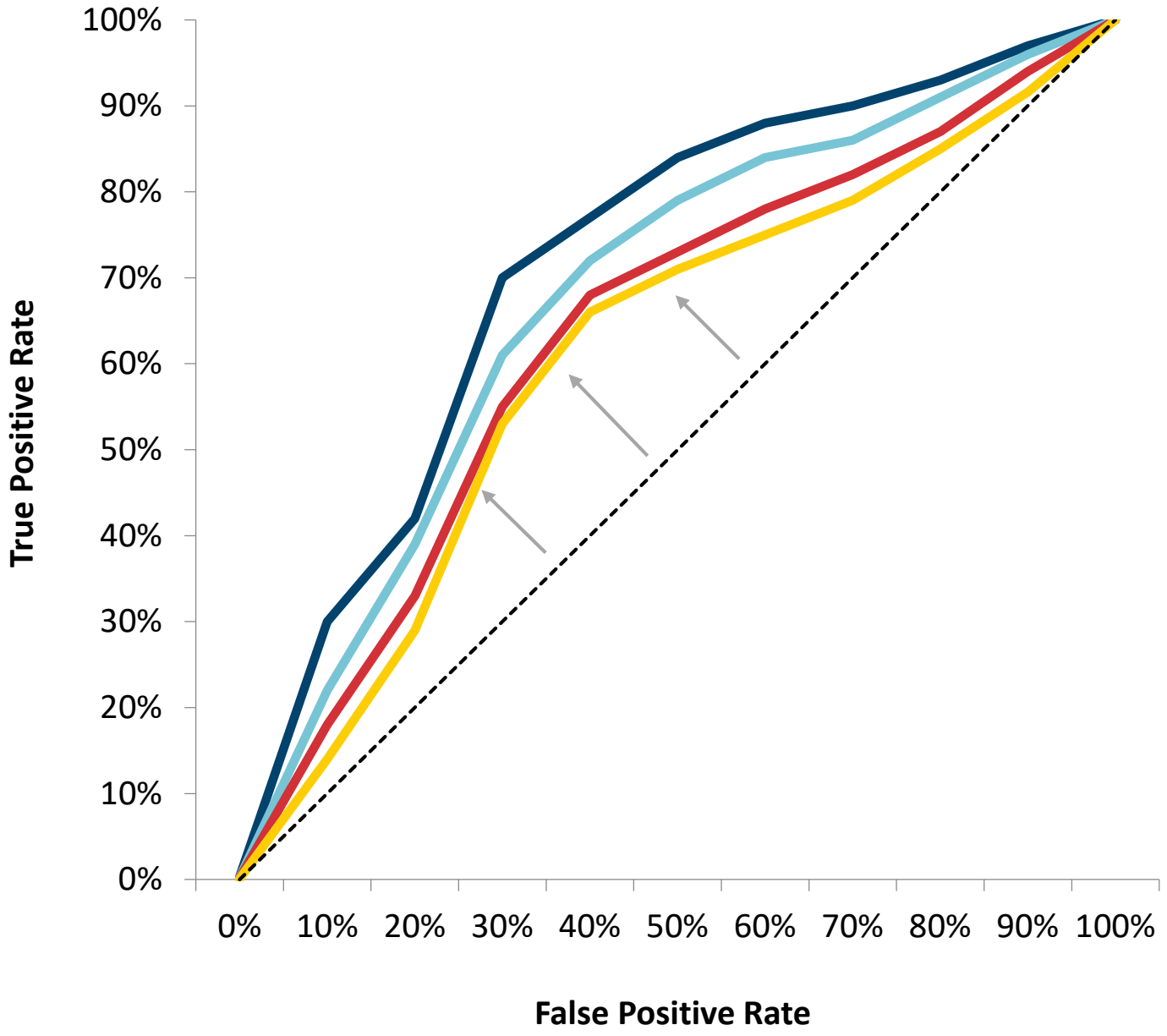

Coefficient Standard Error
0%
2%
4%
6%
8%
10%
Factor
1
2
3
4
5

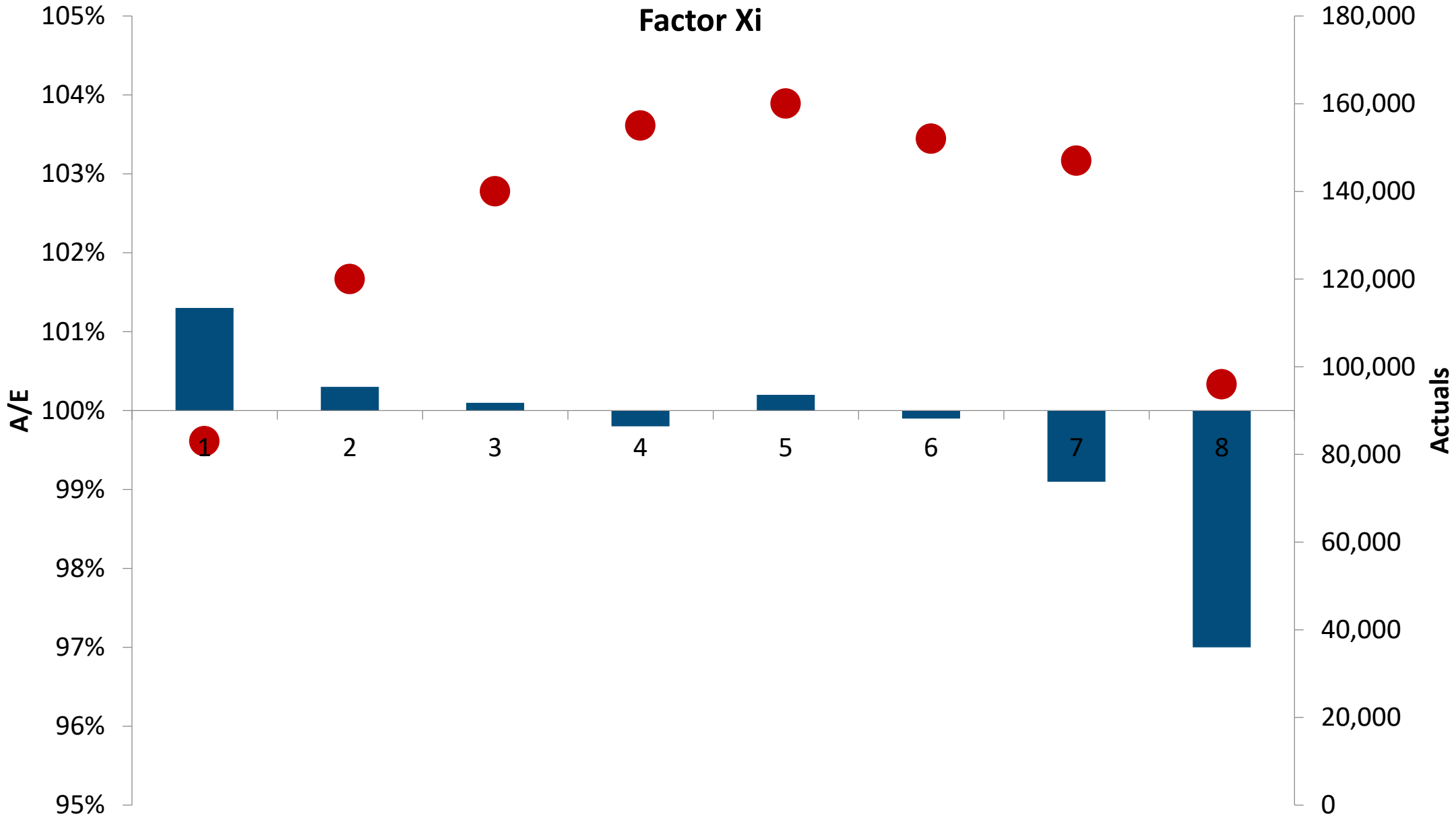
5-Fold Cross Validation

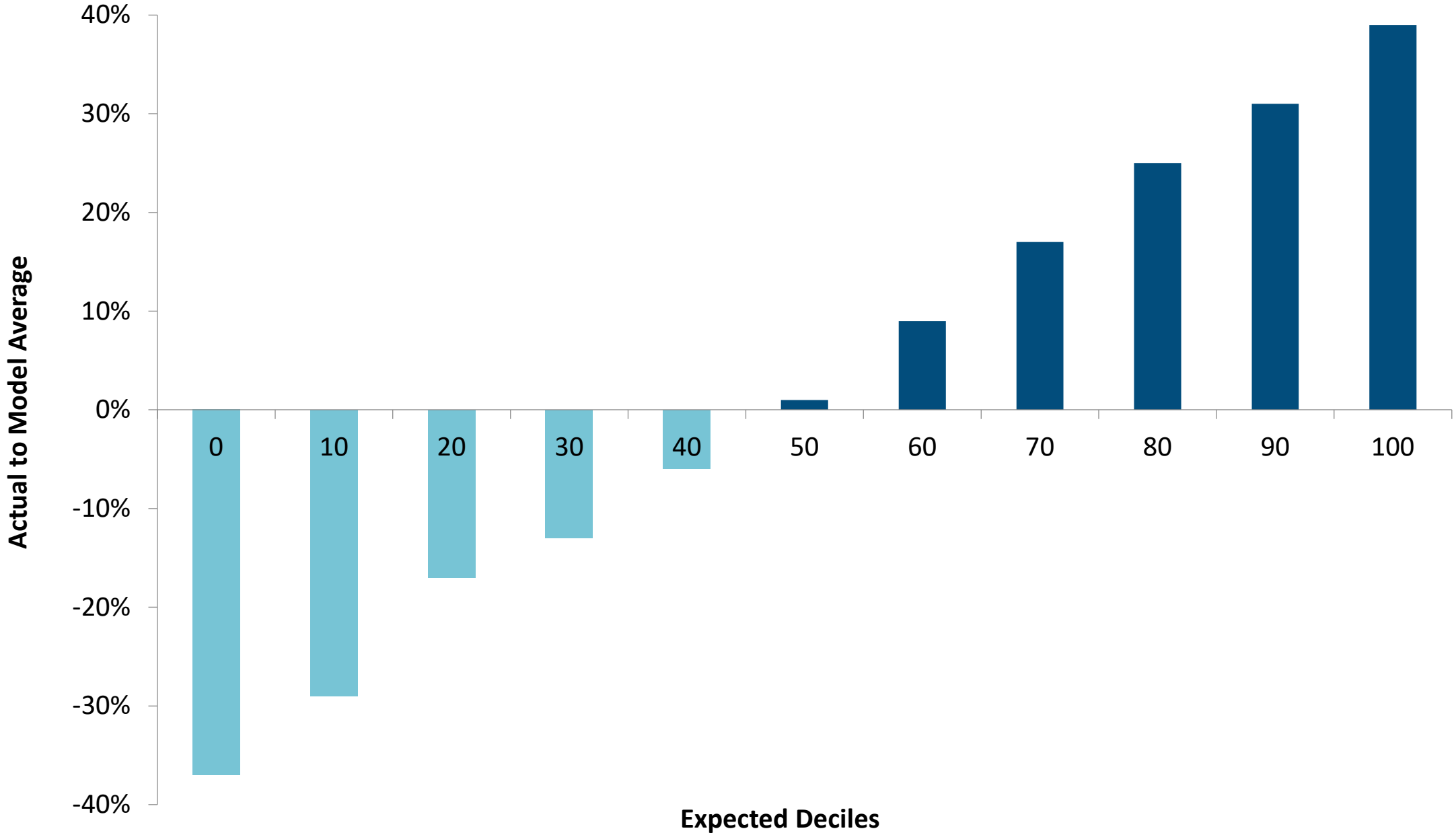
Measures the bias-variance trade-off




3.00%
2.00%
1.00%
0.00%
Avg Abs A/E Error
0
1
2
3
4
5
6
7
8
9
10
Number of Factors


True Positive Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
False Positive Rate


Factor Xi
105%
104%
103%
102%
101%
100%
99%
98%
97%
96%
95%
A/E
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
Actuals
1
2
3
4
5
6
7
8


40%
30%
20%
10%
0%
-10%
-20%
-30%
-40%
Actual to Model Average
0
10
20
30
40
50
60
70
80
90
100
Expected Deciles

Improving Models with Industry Data

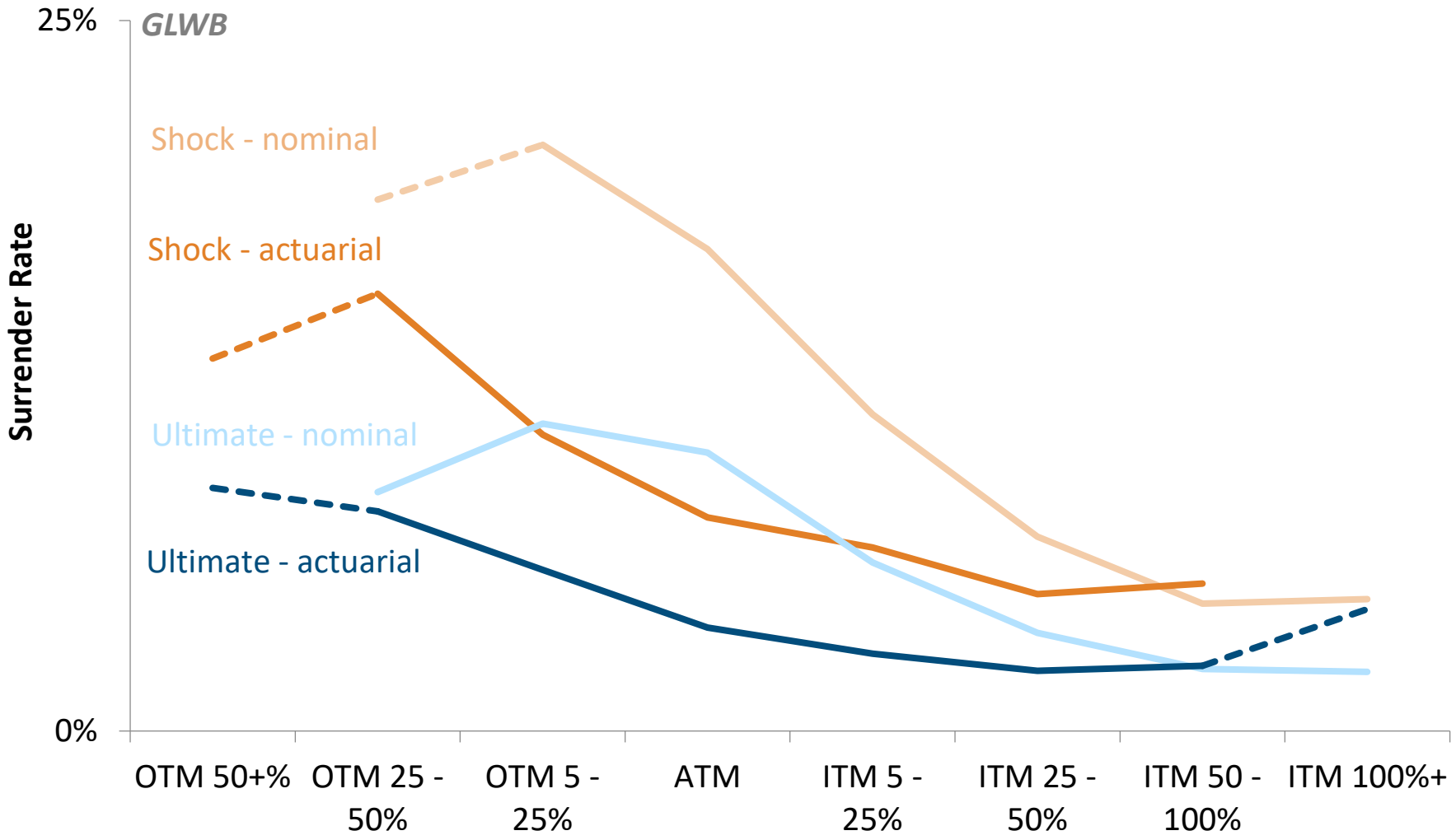
Results vary over time and between companies

- Each company's size affects quality of analytical insights and volatility of their own results (a credibility problem)
- Obvious composition differences
- Subtler idiosyncratic differences (product feature nuances, distribution channels, operational practices, open/closed blocks, etc)
- Using only your data, it is very difficult to identify the signal from the noise

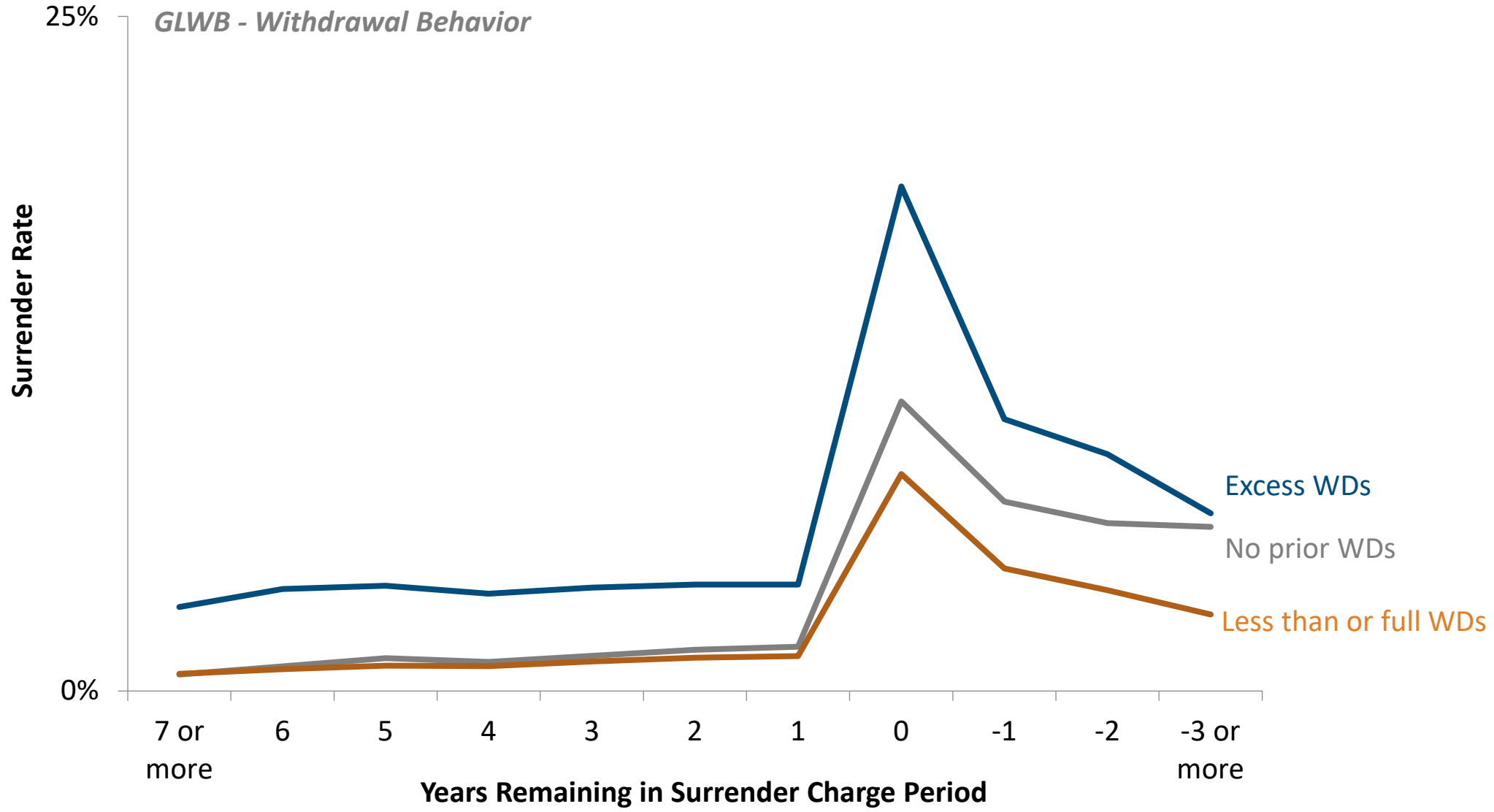
Variable annuity industry data

- 24 companies
- Seriatim monthly data for policyholder behavior and mortality
- January 2008 through December 2018
- $795 billion ending account value

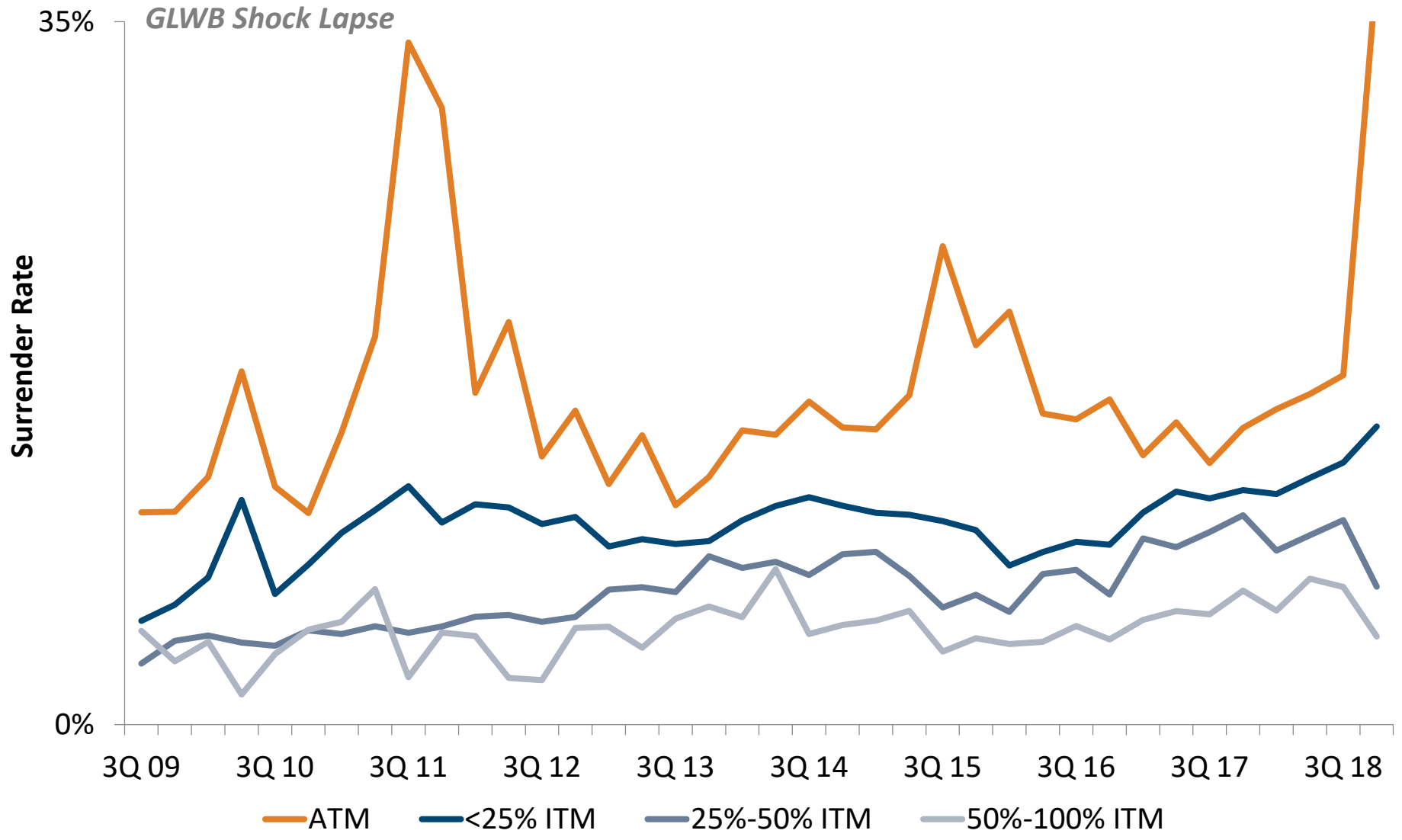
How you measure value matters, and credibility is vastly improved with industry data



Industry data shows that surrender rates are lower when income features are utilized…

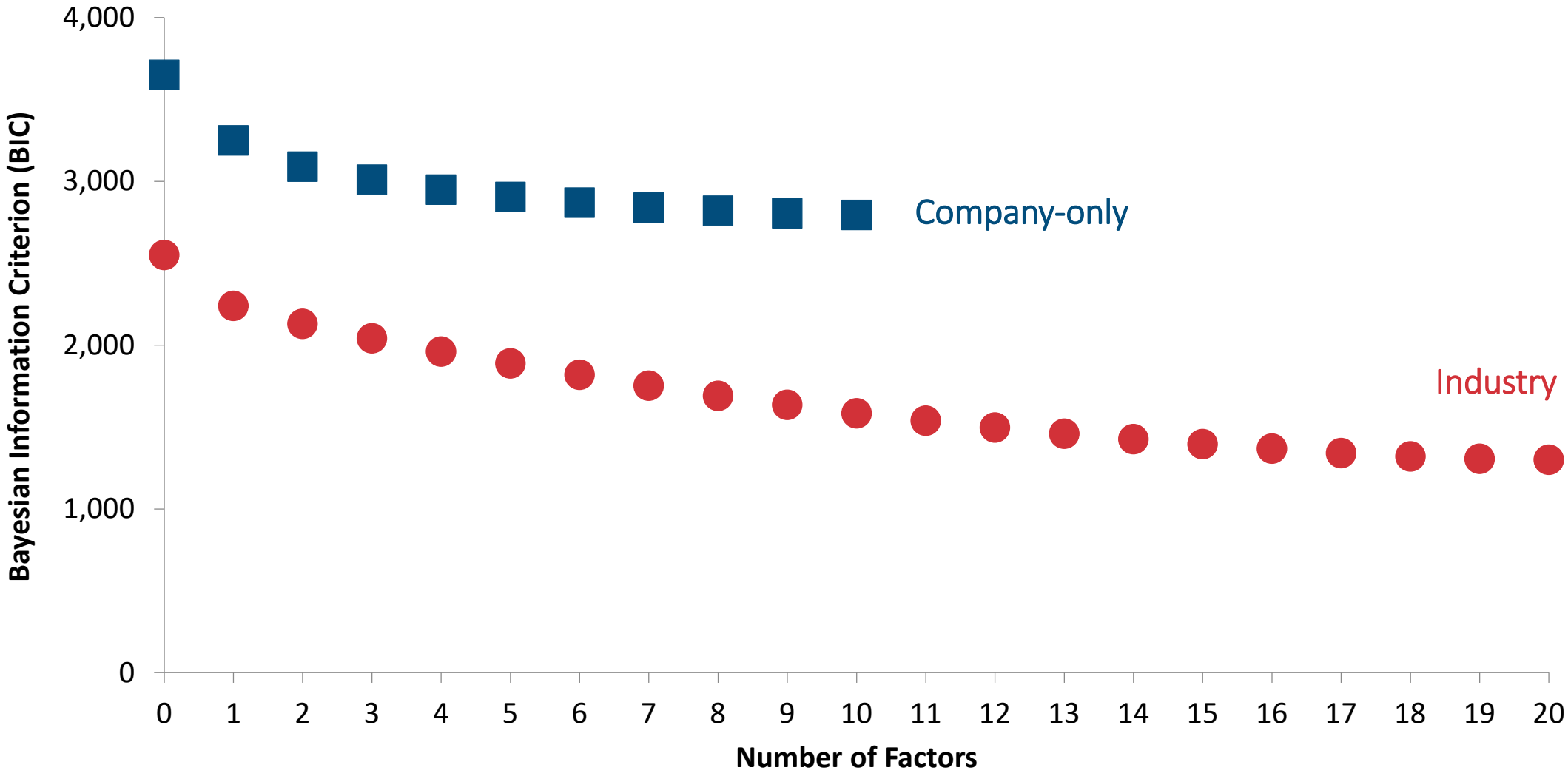


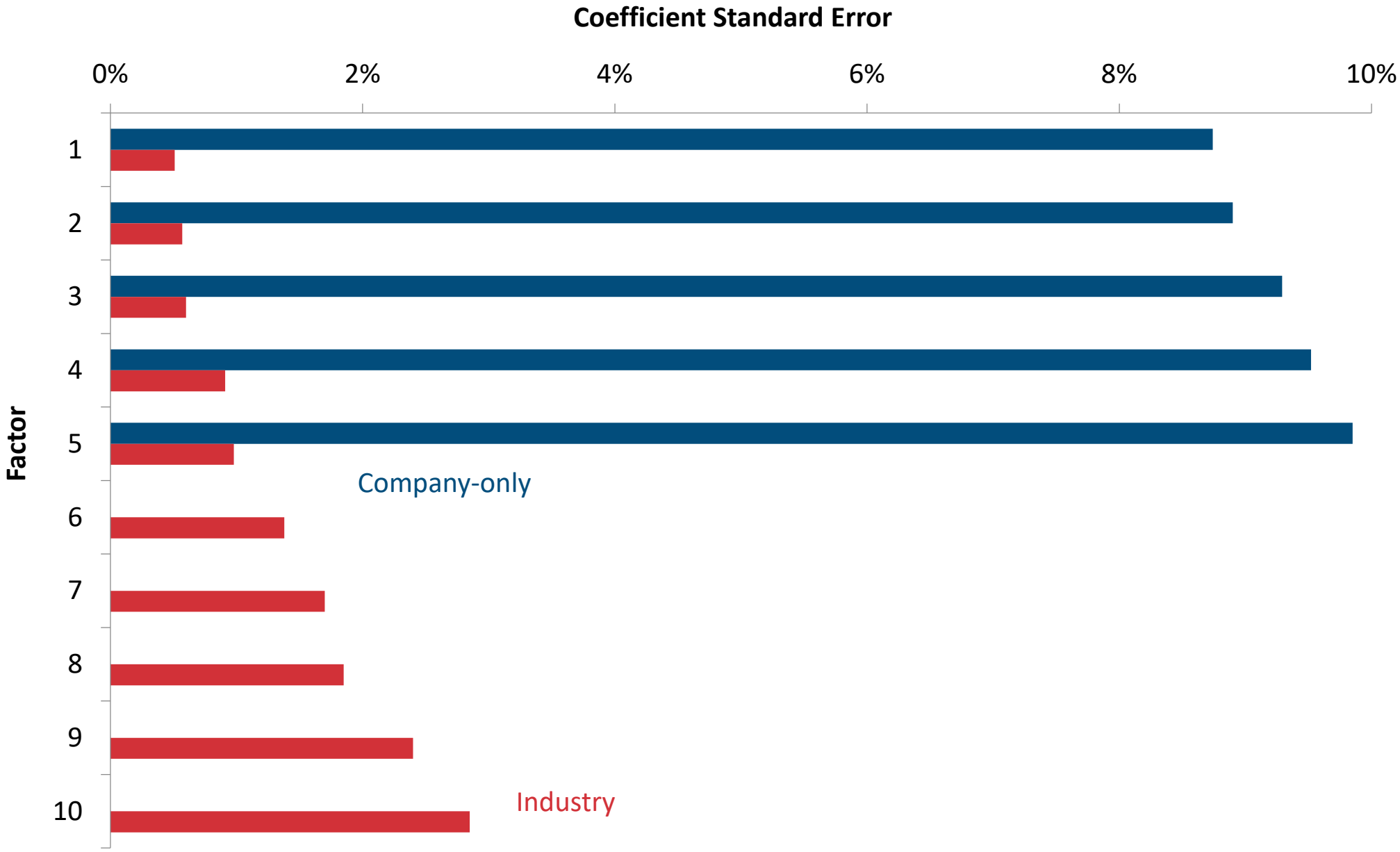
…and dynamic lapse sensitivity varies

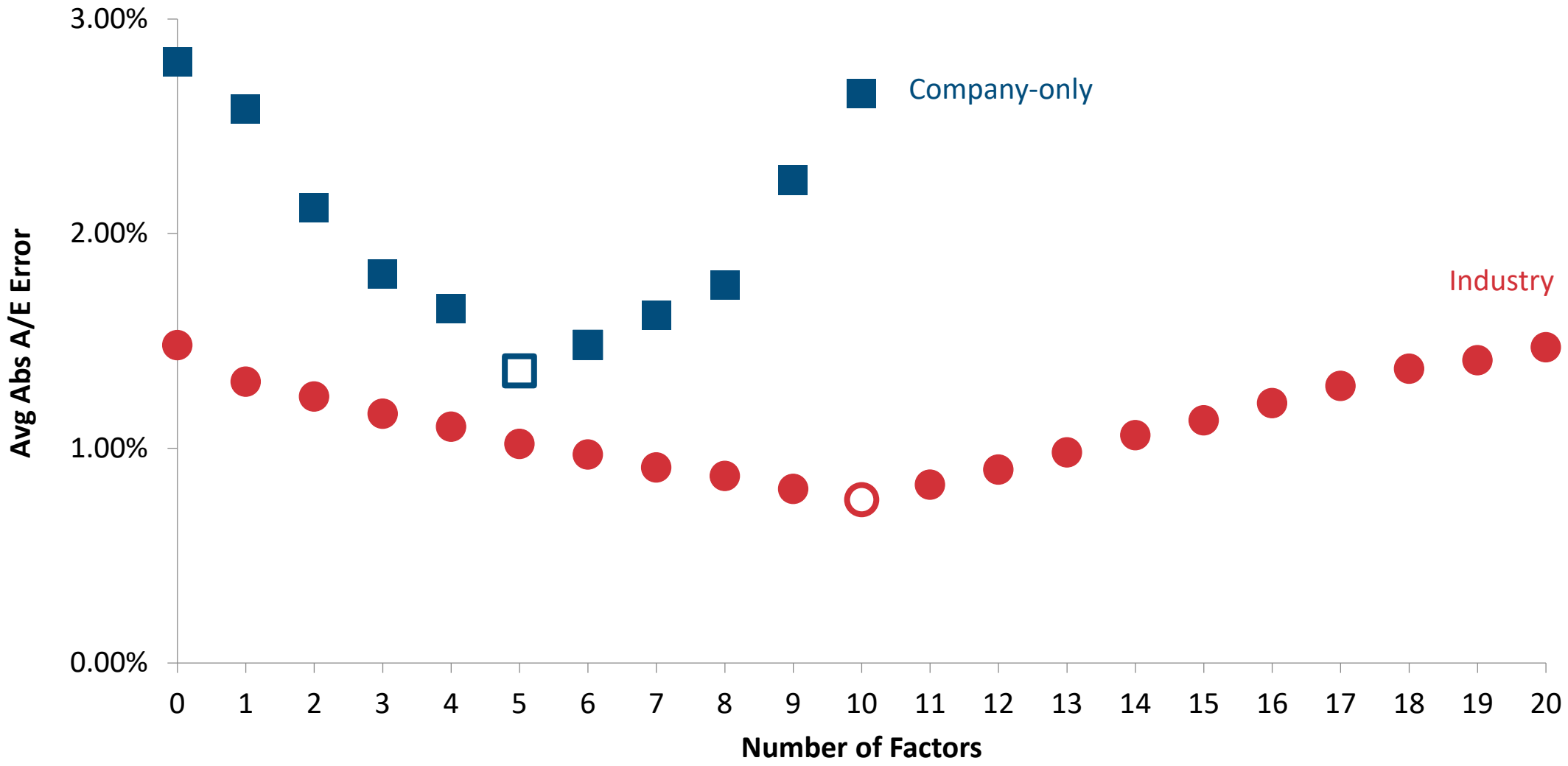


Modeling and assumptions

- Measuring goodness-of-fit for candidate models
- Testing predictive power on out-of-sample data
- ➢ Using relevant industry data to improve candidate models
- Art + science: choosing, communicating, and ongoing recalibration

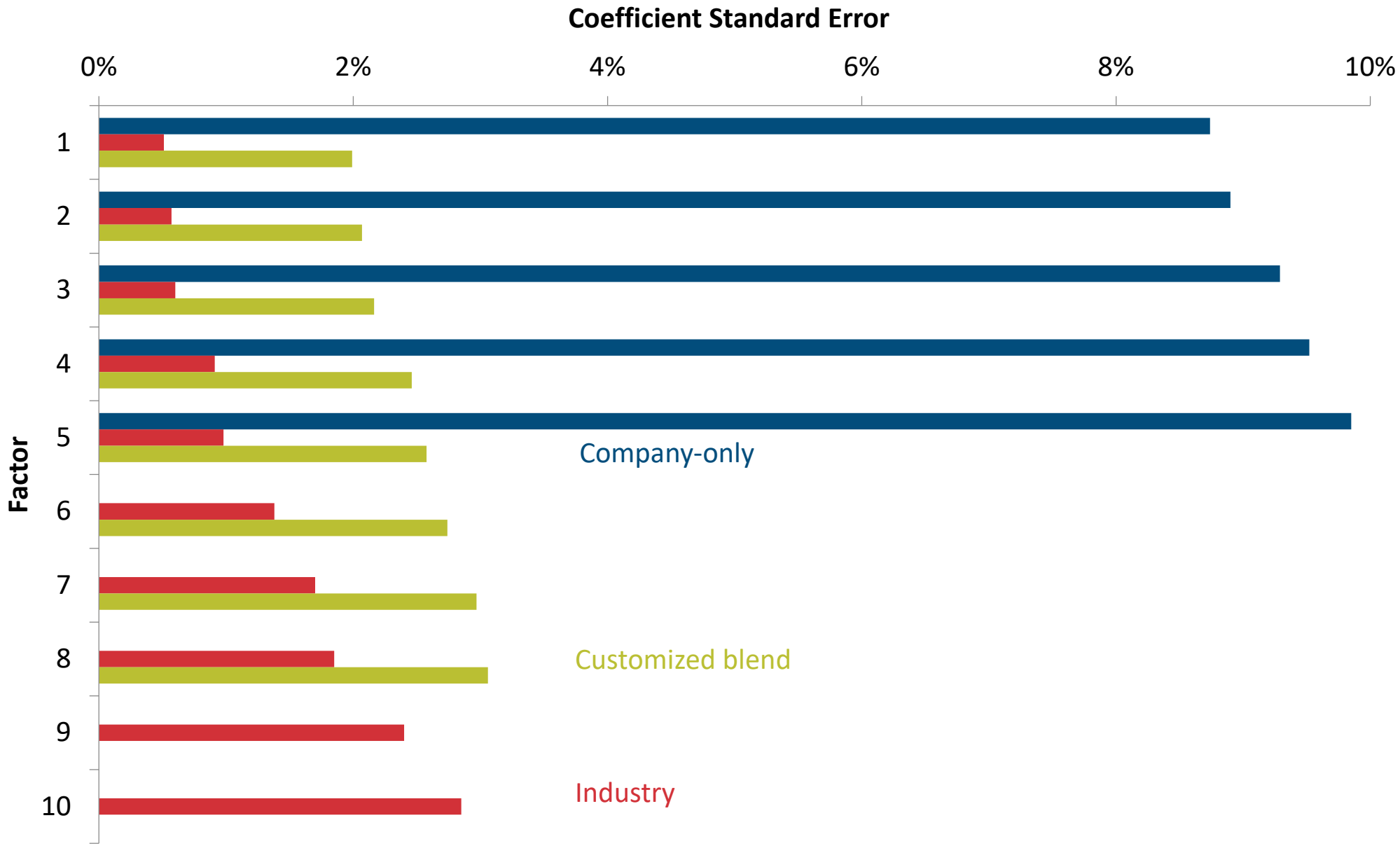

4,000
3,000
2,000
1,000
0
Bayesian Information Criterion (BIC)
Company-only
Industry
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Number of Factors

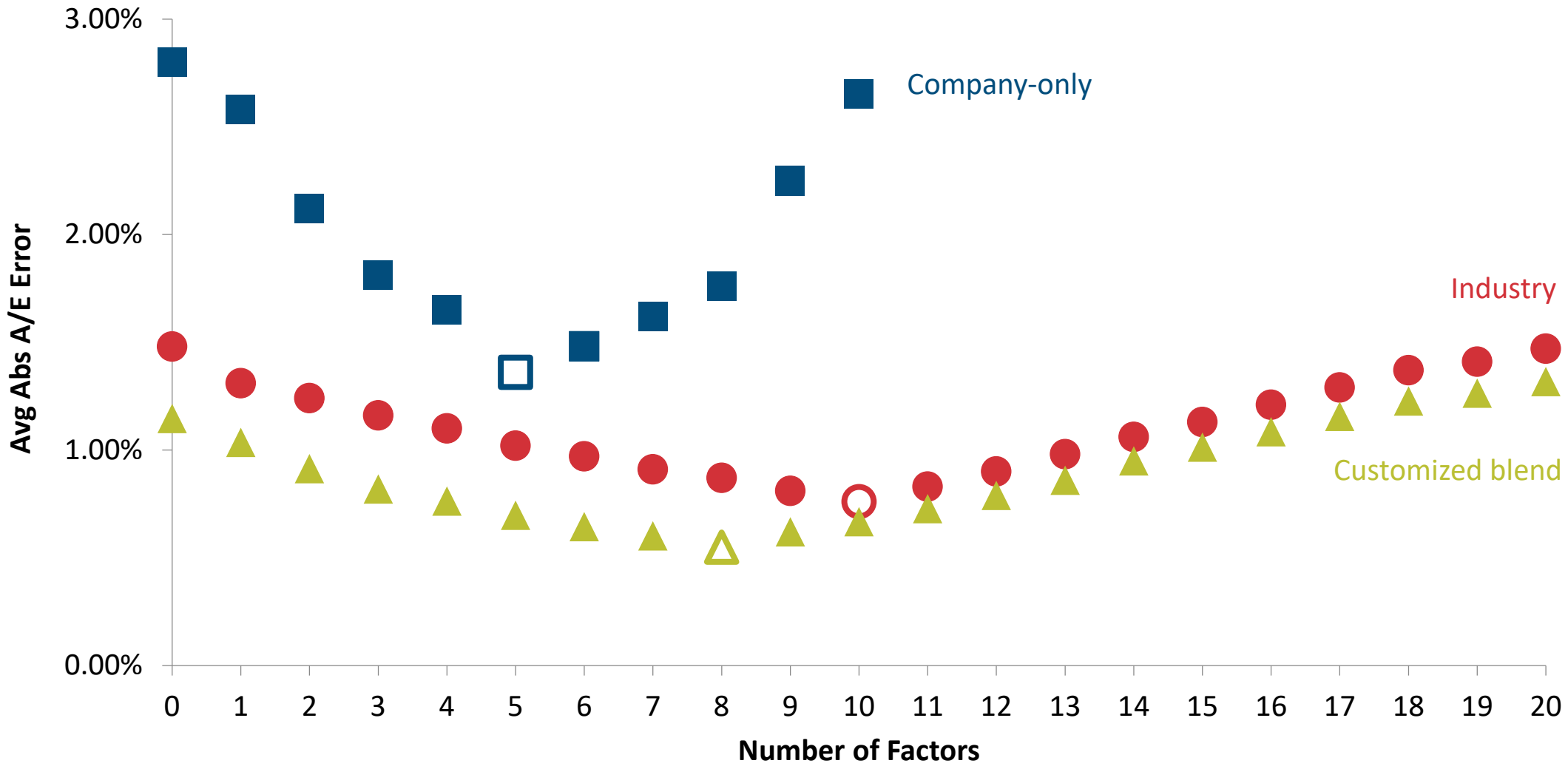

Coefficient Standard Error
0%
2%
4%
6%
8%
10%
Factor
1
2
3
4
5
6
7
8
9
10
Company-only
Industry


3.00%
2.00%
1.00%
0.00%
Avg Abs A/E Error
Company-only
Industry
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Number of Factors

Customize your model in a credibility-based framework

- Subject matter expertise
- Actuarial judgment
- Quantify the benefits of using relevant industry data
- Ongoing recalibration, so focus on the framework

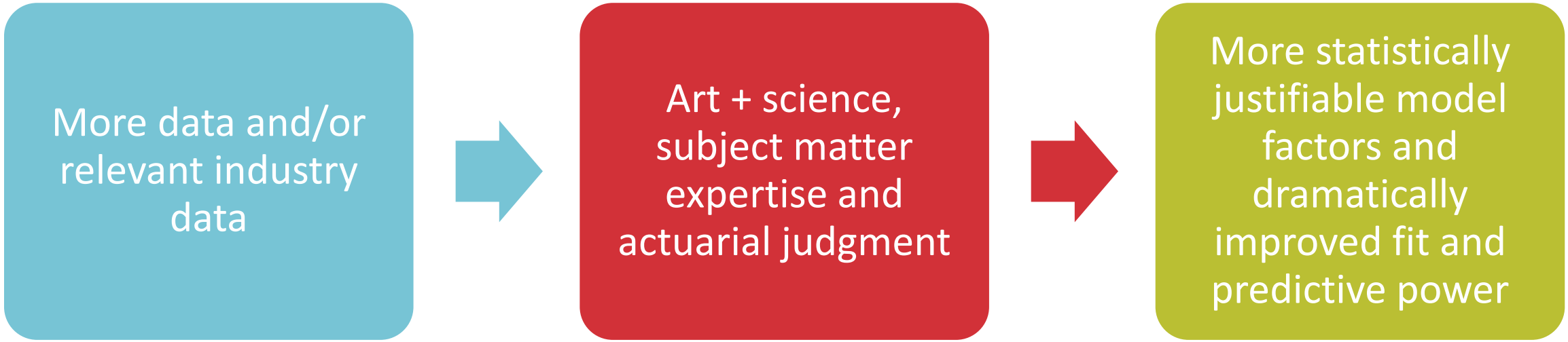

Coefficient Standard Error
0%
2%
4%
6%
8%
10%
Factor
1
2
3
4
5
6
7
8
9
10
Company-only
Customized blend
Industry


3.00%
2.00%
1.00%
0.00%
Avg Abs A/E Error
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Number of Factors
Company-only
Industry
Customized blend

Improving models with industry data

- Customize your model in a credibility-based framework
- Quantify the improvement in goodness-of-fit and predictive power metrics
- Quantify these improvements in financial terms
- Quantify the cost to access and use relevant industry data
- Altogether, does this improve your financial risk profile?

Learnings


More data and/or relevant industry data
Art + science, subject matter expertise and actuarial judgment
More statistically justifiable model factors and dramatically improved fit and predictive power


SOCIETY OF
ACTUARIES®