56 - Avoiding Statistical Pitfalls in Actuarial Work

TS 56: Avoiding Statistical Pitfalls in Actuarial Work

Douglas L. Robbins

August 27, 2019

PACIFIC LIFE

Topics

- Tail Scenarios and “Sensitivity Dependency”
- CTE 98 Challenges
- Use and Misuse of Linear Regression

Topics

- Tail Scenarios and “Sensitivity Dependency”

Sensitivity Testing in General

- For any Risk Factor, in whatever situation
  - You may want to examine the impact of stressing that assumption on your CTE, or other, testing results
  - You may have an idea of what “moderately adverse” all the way to “worst case” looks like for that assumption
  - You may think that this gives you a feel for the impact on Reserves and/or Capital, if that assumption goes south
- Your technique may be very statistical – which is both the good news and the bad . . . .

Something to Consider . . . .

- It's not necessarily true (or even likely) that your risks are independent, so how do you combine your various risk factors?
  - Some risk factors, like economy and lapse, may be "prima facie ***dependent***"
  - Others, like lapse and basic expenses might seem to be independent, but in the end not be
- A couple simple examples to try to make my meaning clear:

The Ice Cream / Oreo Paradox

- Consider a proposed relationship between outside temperature, and how much ice cream you want (Sm, Md, Lg)
  - I would claim this is “prima facie dependent”
  - The hotter it is, the more ice cream you want!
- But . . . what if you are considering your desire for ice cream, alongside your desire for Oreo cookies? Independent?
- Let’s assume so; but what if . . . .

What If . . . .

- . . . . you know your spouse will kill you if you eat a whole carton & a whole box, in one day?
- How effectively will you analyze the "Risk Margin of your death" if you only hold one "snack" at "Medium" and stress the other?
- You won't – you will die!

- Moral: Two independent events can become dependent in terms of tail risk
- Examining sensitivities in tandem may be wise

Topics

- CTE 98 Challenges

Imminent Changes for Variable Annuities

- You've probably heard about many of them throughout this meeting

Imminent Changes for Variable Annuities

- You've probably heard about many of them throughout this meeting
- The key one I will focus on here is the use of CTE 98 for VA C-3, as opposed to a related derivative of CTE 90

Imminent Changes for Variable Annuities

- You've probably heard about many of them throughout this meeting
- The key one I will focus on here is the use of CTE 98 for VA C-3, as opposed to a related derivative of CTE 90
- The first difference between the two that may occur to you is sample size – for a given # of overall scenarios, this will pick up 1/5 as many tail scenarios – is that ok?

Imminent Changes for Variable Annuities

- You've probably heard about many of them throughout this meeting
- The key one I will focus on here is the use of CTE 98 for VA C-3, as opposed to a related derivative of CTE 90
- The first difference between the two that may occur to you is sample size – for a given # of overall scenarios, this will pick up 1/5 as many tail scenarios – is that ok? What sample size are we taught is "enough"?

The E-Lottery

- I’m going to continue this segment by telling you about an online lottery I’ve always wanted to form

The E-Lottery

- I’m going to continue this segment by telling you about an online lottery I’ve always wanted to form
- After I explain why, you’ll think I’m even more of a geek – that’s ok!

The E-Lottery

- I’m going to continue this segment by telling you about an online lottery I’ve always wanted to form
- After I explain why, you’ll think I’m even more of a geek – that’s ok!
- In this lottery, I find a large number (“N”) of players, who give me $1 each

The E-Lottery

- I'm going to continue this segment by telling you about an online lottery I've always wanted to form
- After I explain why, you'll think I'm even more of a geek – that's ok!
- In this lottery, I find a large number ("N") of players, who give me $1 each

- Each picks a number between 1 and N, and I eventually will randomly select a ball from a basket with balls marked with 1 -> N

The E-Lottery

- The players all make their own choices, and cannot collaborate (so “independent”)
- If one player alone picks the right number, they win the entire pot of $N
- If multiple players win, they split $N evenly
- If no one picks the right number, I keep it all
- What are the odds that I keep it all?

The E-Lottery

- In fact, each player has a 1/N chance of hitting the right number
- If there's at least one winner, I lose (unless it's a buddy!) – I get nothing
- So what I want is $(1 - 1/N)^N$

The E-Lottery

- In fact, each player has a 1/N chance of hitting the right number
- If there's at least one winner, I lose (unless it's a buddy!) – I get nothing
- So what I want is $(1 - 1/N)^N$ as N -> infinity!

The E-Lottery

- In fact, each player has a 1/N chance of hitting the right number
- If there's at least one winner, I lose (unless it's a buddy!) – I get nothing
- So what I want is $(1 - 1/N)^N$ as N -> infinity!
- Anyone recognize that formula?

The E-Lottery

- In fact, each player has a 1/N chance of hitting the right number
- If there's at least one winner, I lose (unless it's a buddy!) – I get nothing
- So what I want is $(1 - 1/N)^N$ as N -> infinity!
- Anyone recognize that formula?
- That's right!! The answer is 1/e!!

The E-Lottery

- In fact, each player has a 1/N chance of hitting the right number
- If there's at least one winner, I lose (unless it's a buddy!) – I get nothing
- So what I want is $(1 - 1/N)^N$ as N -> infinity!
- Anyone recognize that formula?
- That's right!!   The answer is 1/e!!
- Was I right about the "geek" thing?

WHAT did we just spend all that time on?

- This concept actually matters, as it makes us a bit more intuitive about tail samples of highly-skewed distributions
- If I run 100 scenarios, for instance, what are my odds that none of them exceed the 99^{th} %-ile worst true result?
- Very close to 1/e, which is A LOT, in terms of a statistical risk!!

You'd never do this, but as a thought exercise:

- What if I ran 100 scenarios to do my VACARVM and C-3, II studies?
- Under a CTE 90 regime, is that 1/e result important? Sure, somewhat
- But how about a CTE 98 regime?
  - I'm basically going to look at 2 scenarios, and I'm only about {1-1/e} confident that at least one of them is greater than the true 99th percentile (CTE's mid-point)!!
  - It should be apparent that this estimator is badly biased downward—a deeply unconservative outcome

Downward bias is a “thing” with any CTE

- Another way to make this intuitively obvious is to consider the nature of the “worst result”
- It is always your most-biased estimate of any %-ile in your study

Downward bias is a "thing" with any CTE

- Another way to make this intuitively obvious is to consider the nature of the "worst result"
- It is always your most-biased estimate of any %-ile in your study
  - Each new observation you record can only do 1 of 2 things

Downward bias is a “thing” with any CTE

- Another way to make this intuitively obvious is to consider the nature of the “worst result”
- It is always your most-biased estimate of any %-ile in your study
  - Each new observation you record can only do 1 of 2 things
  - Either your previous worst-case stays the same, or the new observation makes your estimate even worse

Downward bias is a “thing” with any CTE

- Another way to make this intuitively obvious is to consider the nature of the “worst result”
- It is always your most-biased estimate of any %-ile in your study
  - Each new observation you record can only do 1 of 2 things
  - Either your previous worst-case stays the same, or the new observation makes your estimate even worse
  - This effect ripples through your entire set of CTE observations, although it quickly becomes less material

Downward bias is a "thing" with any CTE

- Another way to make this intuitively obvious is to consider the nature of the "worst result"
- It is always your most-biased estimate of any %-ile in your study
  - Each new observation you record can only do 1 of 2 things
  - Either your previous worst-case stays the same, or the new observation makes your estimate even worse
  - This effect ripples through your entire set of CTE observations, although it quickly becomes less material
- Still, you can see even with 1,000 scenarios, the jump from CTE 90 to 98 is a big deal

How Big a Deal is it?

- Without knowing your loss distribution, I can't be precise

- However, I'd guess most capital runs produce many PV(losses) = $0, followed by a set of rapidly escalating positives

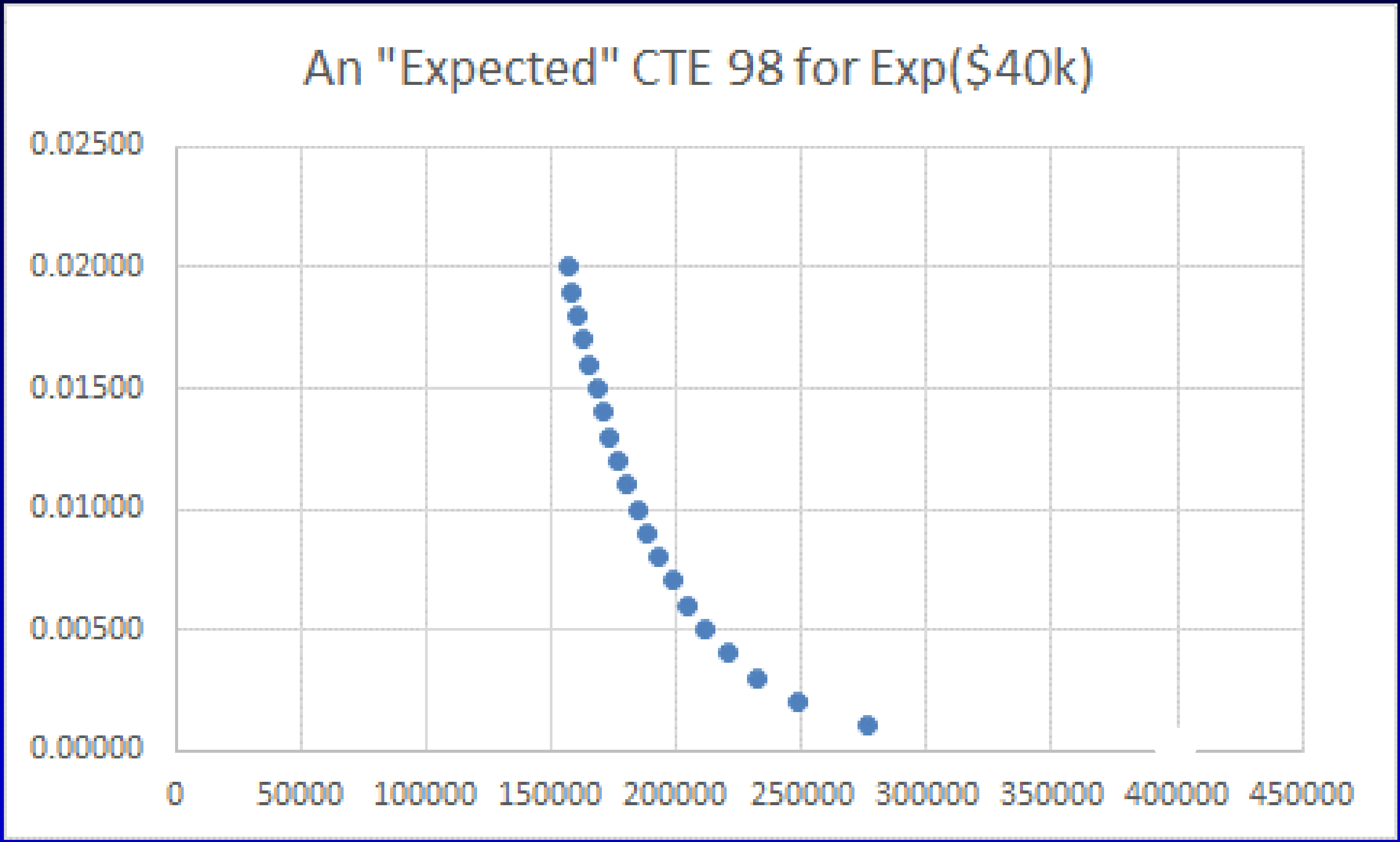
- Plausibly more or less Exponential

- Let's assume that, and see what happens . . . .

How Big a Deal is it?

- I'm going to steal a parameter estimate from a VA WB study I did recently
  - Total Premium around $8.7 million
  - Mean and Std Dev around $40k, so we'll use that for the Exponential lambda
- Now a WB loss distribution is actually bounded (at least if mortality -> 1.000)
- So my real distribution is probably less skewed than Exponential, but if we assume Exp, we can talk "true parameter values"

How Big a Deal is it?



A few key facts about EXP($40k) and CTE98

- My true 98^{th} percentile is about $157k
  - P(981^{st} ranked scenario >= 98^{th} %-ile) = ?

A few key facts about EXP($40k) and CTE98

- My true 98^{th} percentile is about $157k
  - P(981st ranked scenario >= 98^{th} %-ile) = ?
  - It's only about 62%

A few key facts about EXP($40k) and CTE98

- My true 98^{th} percentile is about $157k
  - P(981^{st} ranked scenario >= 98^{th} %-ile) = ?
  - It's only about 62%
- My true 99.9^{th} %-ile is about $276k

A few key facts about EXP($40k) and CTE98

- My true 98^{th} percentile is about $157k
  - P(981^{st} ranked scenario >= 98^{th} %-ile) = ?
  - It's only about 62%
- My true 99.9^{th} %-ile is about $276k
  - P($1,000^{th}$ ranked scenario >= 99.9^{th} %-ile) = ?

A few key facts about EXP($40k) and CTE98

- My true 98th percentile is about $157k
  - P(981st ranked scenario >= 98th %-ile) = ?
  - It's only about 62%
- My true 99.9th %-ile is about $276k
  - P(1,000th ranked scenario >= 99.9th %-ile) = ?
  - Hint: You should know this one!

A few key facts about EXP($40k) and CTE98

- My true 98^{th} percentile is about $157k
  - P(981st ranked scenario >= 98^{th} %-ile) = ?
  - It's only about 62%
- My true 99.9^{th} %-ile is abo
  - P(1,000th ranked scenario >= 99.9^{th}
  - Hint: You should know this one!

A few key facts about EXP($40k) and CTE98

- My true 98th percentile is about $157k
  - P(981st ranked scenario >= 98th %-ile) = ?
  - It's only about 62%
- My true 99.9th %-ile is about $276k
  - P(1,000th ranked scenario >= 99.9th %-ile) = ?
  - Hint: You should know this one!
  - Yes, that's right! It's 1 - 1/e, or about 63%

A few key facts about EXP($40k) and CTE98

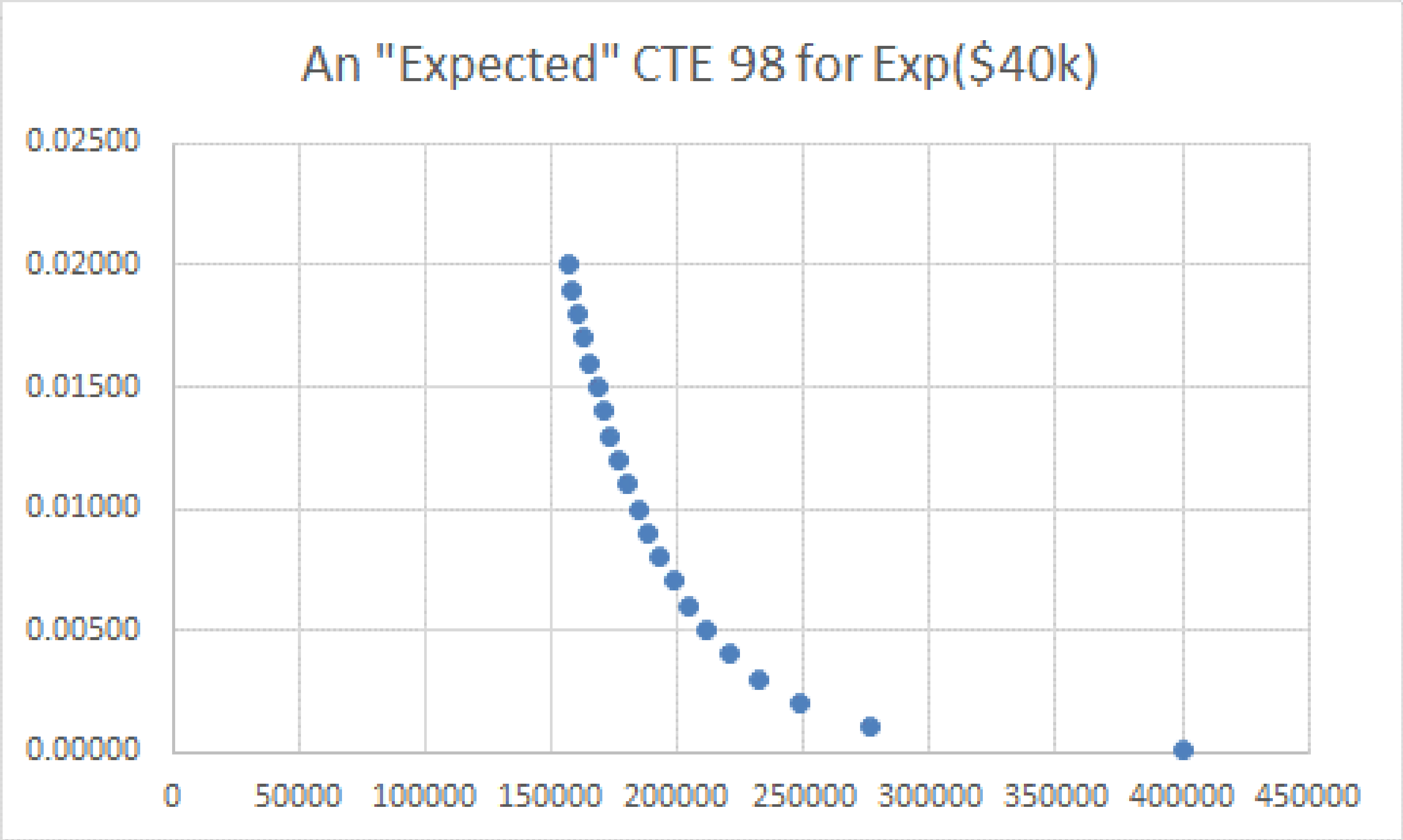
- My true 98^{th} percentile is about $157k
  - P(981^{st} ranked scenario >= 98^{th} %-ile) = ?
  - It's only about 62%
- My true 99.9^{th} %-ile is about $276k
  - P($1,000^{th}$ ranked scenario >= 99.9^{th} %-ile) = ?
  - Hint: You should know this one!
  - Yes, that's right! It's 1 - 1/e, or about 63%
- The two are probably correlated due to a sort of accordion/spring effect (if you think about it); leading us to . . .

A few key facts about EXP($40k) and CTE98

- The difference between a CTE 98 that runs evenly from 98th to 99.9th percentiles over 20 tail scenarios, and one that runs from 97.9th to 99.8th, is about $6k
- The std error of the CTE98 is about $10k
- That means the bias is quite relevant
- But even that fails to account for the fact that a CTE98 should run from the 98th to the 100th percentile (not the 99.9th)

A few key facts about EXP($40k) and CTE98

- I actually deleted something from the chart I showed you before; the real chart is:



A few key facts about EXP($40k) and CTE98

- The "true" value for the 100^{th} percentile is of course, infinity – can't do much with that
- But if we consider the 98^{th} through the 99.99^{th} percentile range as our true CTE, the potential bias in our estimate is more like $15k, which is around {std error X 150%}
- This is likely unacceptable
- If we run 10,000 scenarios, chances are the bias (not to mention the std error) will be small enough to safely ignore

Bottom Line Lesson

- When working out a CTE98 . . .

Bottom Line Lesson

- When working out a CTE98 . . .
- . . . just always err on the side of using MORE SCENARIOS

Bottom Line Lesson

- When working out a CTE98 . . .
- . . . just always err on the side of using MORE SCENARIOS

Bottom Line Lesson

- When working out a CTE98 . . .

- . . . just always err on the side of using MORE SCENARIOS

- Or perhaps . . . just perhaps . . . consider using a scenario selection algorithm so you only end up running the tail ones from a hypothetical "large enough" set

Topics

- Use and Misuse of Linear Regression

Why do we do Multiple Linear Regressions?

- We want as accurate a predictor of a future outcome as we can get from our data set

- We instinctively feel that the more data we can harness, the better
  - Sometimes, this is the case—for a Balanced Fund mapping, you may need 3 Equity classes plus Bond
  - But sometimes

Why do we do Multiple Linear Regressions?

- We want as accurate a predictor of a future outcome as we can get from our data set
- We instinctively feel that the more data we can harness, the better
  - Sometimes, this is the case—for a Balanced Fund mapping, you may need 3 Equity classes plus Bond
  - But sometimes . . . . . . well

Why do we do Multiple Linear Regressions?

- We want as accurate a predictor of a future outcome as we can get from our data set
- We instinctively feel that the more data we can harness, the better
  - Sometimes, this is the case—for a Balanced Fund mapping, you may need 3 Equity classes plus Bond
  - But sometimes . . . . . . well . . . . . . let’s look at another Case Study!

Linear Regression Case Study

- An actuary who is nearing retirement has attended Val Act for years and years
- He’s found that when he’s away from home, he can’t track his net worth daily, but he does have access to the S&P
- He has the data going back 9 years of his pre- and post- Val Act data, and wants to be able to predict his 2019 end-of-meeting net worth before going home

Linear Regression Case Study

- Here is what he knows:
  - His basket of stocks, currently worth $100k, tracks at least somewhat with the S&P 500, but not perfectly
  - He doesn’t always like the Val Act food
- Here is what WE KNOW:
  - His basket of stocks actually moves proportionately with the S&P, with an error that is Uniform(+/- 5%)
  - If he eats the Val Act food, he spends nothing extra, but:
    - If he goes for a street hotdog, he’ll spend U($5-$10)
    - If he goes out to “Sushi Max,” he’ll spend U($25-50)

Linear Regression Case Study

- Again, this actuary has 9 years of prior data; scaling the initial portfolio value each year to the current $100k, the history is:

	S&P -10%	S&P Flat	S&P +10%
Meeting Meal	$89,133	$96,833	$110,720
Hot Dog	$91,736	$104,641	$107,137
Sushi Max	$90,114	$100,029	$112,317

- This seems eminently “regressable” so the actuary goes to it! I mean, why not?

Linear Regression Case Study

- The actuary, sensibly, first looks at S&P alone:
  - X_1-data are -1 -> -10%, 0, 1 -> +10% (just to clarify)
  - The R^2 is 96.1% and the p-value is 4 * 10^-5
  - The parameter on the S&P500 is $9,865 with standard error $1,075—so far, so good
- They then look at food (X_2 data 0-2):
  - The R^2 is 9.4% and the p-value is 81%--shocking, right?
  - The parameter on food is $962 with standard error $3,863—so the actual value is within the conf interval!
- The actuary, finally, looks at both in tandem:
- What do you think happened?

Linear Regression Case Study

- NOT WHAT I THOUGHT WOULD HAPPEN!

Linear Regression Case Study

- NOT WHAT I THOUGHT WOULD HAPPEN!
- None of us probably had enough stats in school, and what we did store up leaks over time

Linear Regression Case Study

- NOT WHAT I THOUGHT WOULD HAPPEN!
- None of us probably had enough stats in school, and what we did store up leaks over time
- But the result makes total sense:

Linear Regression Case Study

- NOT WHAT I THOUGHT WOULD HAPPEN!
- None of us probably had enough stats in school, and what we did store up leaks over time
- But the result makes total sense:
  - The R^2 went up to 96.5% (it will increase with each new variable you add, every time)

Linear Regression Case Study

- NOT WHAT I THOUGHT WOULD HAPPEN!
- None of us probably had enough stats in school, and what we did store up leaks over time
- But the result makes total sense:
  - The R^2 went up to 96.5% (it will increase with each new variable you add, every time)
  - The overall F-statistic went up to 0.0003 (so worse than S&P only), with a p-value on S&P of 0.0001

Linear Regression Case Study

- NOT WHAT I THOUGHT WOULD HAPPEN!
- None of us probably had enough stats in school, and what we did store up leaks over time
- But the result makes total sense:
  - The R^2 went up to 96.5% (it will increase with each new variable you add, every time)
  - The overall F-statistic went up to 0.0003 (so worse than S&P only), with a p-value on S&P of 0.0001
  - But . . . the parameter estimates did not change one iota

Linear Regression Case Study

- NOT WHAT I THOUGHT WOULD HAPPEN!
- None of us probably had enough stats in school, and what we did store up leaks over time
- But the result makes total sense:
  - The R^2 went up to 96.5% (it will increase with each new variable you add, every time)
  - The overall F-statistic went up to 0.0003 (so worse than S&P only), with a p-value on S&P of 0.0001
  - But . . . the parameter estimates did not change one iota
  - The reason is that the actuary has designed a perfectly independent set of experimental trials

Lessons so Far

- Nothing too spectacular

Lessons so Far

- Nothing too spectacular
- We all know the probability of a false positive is important

Lessons so Far

- Nothing too spectacular
- We all know the probability of a false positive is important, and we quantify it—what about a false negative?

Lessons so Far

- Nothing too spectacular
- We all know the probability of a false positive is important, and we quantify it—what about a false negative?
- Often we have no clue what that probability is

Lessons so Far

- Nothing too spectacular
- We all know the probability of a false positive is important, and we quantify it—what about a false negative?
- Often we have no clue what that probability is
- One time it's big is when a variable's range is dwarfed by a regression's overall error

What if the Actuary is more like Us?

- On a down market day, they would never go for sushi? And on an up market day, they would never eat Val Act food?
- We go from this old data set:

	S&P -10%	S&P Flat	S&P +10%
Meeting Meal	$89,133	$96,833	$110,720
Hot Dog	$91,736	$104,641	$107,137
Sushi Max	$90,114	$100,029	$112,317

What if the Actuary is more like Us?

- On a down market day, they would never go for sushi? And on an up market day, they would never eat Val Act food?

- To this new data set (same random #s):

		S&P -10%		S&P Flat		S&P +10%
Meeting Meal		$89,133		$96,833		
						$110,713
Hot Dog		$91,736		$104,641		$107,137
		$90,147				
Sushi Max				$100,029		$112,317

What if the Actuary is more like Us?

- On a down market day, they would never go for sushi? And on an up market day, they would never eat Val Act food?*

		S&P -10%		S&P Flat		S&P +10%
Meeting Meal		$89,133		$96,833		
						$110,713
Hot Dog		$91,736		$104,641		$107,137
		$90,147				
Sushi Max				$100,029		$112,317

- We repeat the regression

*Same random numbers used

What do you think happened now?

- The S&P-only regression, of course, changed by an immaterial amount

What do you think happened now?

- The S&P-only regression, of course, changed by an immaterial amount
- But the parameter estimate on food-only was now over $6,500—hmmmm……

What do you think happened now?

- The S&P-only regression, of course, changed by an immaterial amount
- But the parameter estimate on food-only was now over $6,500—hmmmm……
- In a joint regression, the parameter on the S&P500 worsened to $9,192 with a not much changed confidence interval; the food estimate was right at $2,000

Let’s go Whole Hog on this!

- This actuary would in fact never spend a dime on food, during a down market
- They’d only go for sushi to celebrate an up market day (and in that case, they would do so 2/3 of the time)
- Now, what do you think will happen?

Old (2nd set) Data

		S&P -10%		S&P Flat		S&P +10%
Meeting Meal		$89,133		$96,833		
						$110,713
Hot Dog		$91,736		$104,641		$107,137
		$90,147				
Sushi Max				$100,029		$112,317

Newest Data

		S&P -10%		S&P Flat		S&P +10%
Meeting Meal		$89,133		$96,833		
		$91,746				
		$90,155				$110,713
Hot Dog				$104,641		
				$100,063		$107,108
Sushi Max						$112,317

Let’s go Whole Hog on this!

- The S&P-only regression remains solid

Let’s go Whole Hog on this!

- The S&P-only regression remains solid
- The parameter on food is nearly $10,000

Let’s go Whole Hog on this!

- The S&P-only regression remains solid
- The parameter on food is nearly $10,000
  - Does that surprise you?

Let's go Whole Hog on this!

- The S&P-only regression remains solid
- The parameter on food is nearly $10,000
  - Does that surprise you?
  - Should it? Who can guess the R^2 of food??

Let’s go Whole Hog on this!

- The S&P-only regression remains solid
- The parameter on food is nearly $10,000
  - Does that surprise you?
  - Should it? Who can guess the R^2 of food?? It’s 88%!!

Let's go Whole Hog on this!

- The S&P-only regression remains solid
- The parameter on food is nearly $10,000
  - Does that surprise you?
  - Should it? Who can guess the R^2 of food?? It's 88%!!
- In the joint regression, the parameter on the S&P is ~$8,100 and on Food, it is ~$2,100—what strikes you about that?
  - A "significant" relationship (S&P p-level still <1%)
  - A more or less USELESS predictor for the future—WHY??
  - That is the magic of co-linearity, brought to life

Let Me Ask One More Thing

Is the actuary in our example:

Let Me Ask One More Thing

Is the actuary in our example:

- This guy??  (Ebeneezer Scrooge)

Let Me Ask One More Thing

Is the actuary in our example:

- This guy??  (Ebeneezer Scrooge)

- Or this one?? 

(Melvin from "As Good as It Gets")

Let Me Ask One More Thing

Is the actuary in our example:

- This guy??  (Ebeneezer Scrooge)
- Or this one?? 

(Melvin from "As Good as It Gets")

<<Kudos to Jack, btw!!>>

Let Me Ask One More Thing

Is the actuary in our example:

- This guy??  (Ebeneezer Scrooge)

- Or this one??  

(Melvin from "As Good as It Gets")

<<Kudos to Jack, btw!!>>

Let Me Ask One More Thing

Is the actuary in our example:

- This guy??  (Ebeneezer Scrooge)
- Or this one??  

(Melvin from "As Good as It Gets")

'Cause it must be Greed or OCD at work here!!

Let Me Ask One More Thing

Is the actuary in our example:

- This guy??  (Ebeneezer Scrooge)

- Or this one??  

(Melvin from "As Good as It Gets")

'Cause it must be Greed or OCD at work here!! For an actuary, likely the latter

Bottom Line Lessons

- It is probably a waste of time, and only possibly a harmless one, to add regression variables that could be significant, but nevertheless immaterial:
  - Harmless iff the variable added is “independent” of others used
  - The problem is that sometimes you don’t know, and what you don’t know can and will hurt you! Immaterial variables can still be very harmful if included
  - We must resist the temptation to “use it because we can”

Bottom Line Lessons

- It is probably a waste of time, and only possibly a harmless one, to add regression variables that could be significant, but nevertheless immaterial:
  - Harmless iff the variable added is “independent” of others used
  - The problem is that sometimes you don’t know, and what you don’t know can and will hurt you! Immaterial variables can still be very harmful if included
  - We must resist the temptation to “use it because we can”
- “But adding it increases my R^2” isn’t even a ghost of an argument for including a variable

Bottom Line Lessons

- It is probably a waste of time, and only possibly a harmless one, to add regression variables that could be significant, but nevertheless immaterial:
  - Harmless iff the variable added is "independent" of others used
  - The problem is that sometimes you don't know, and what you don't know can and will hurt you! Immaterial variables can still be very harmful if included
  - We must resist the temptation to "use it because we can"
- "But adding it increases my R^2" isn't even a ghost of an argument for including a variable
- Possible Application: Including "Cash" in a fund-mapping exercise

AVOIDING STATISTICAL PITFALLS IN ACTUARIAL WORK

AUGUST 27, 2019

Mark Spong, FSA, CERA, MAAA
Senior Consultant

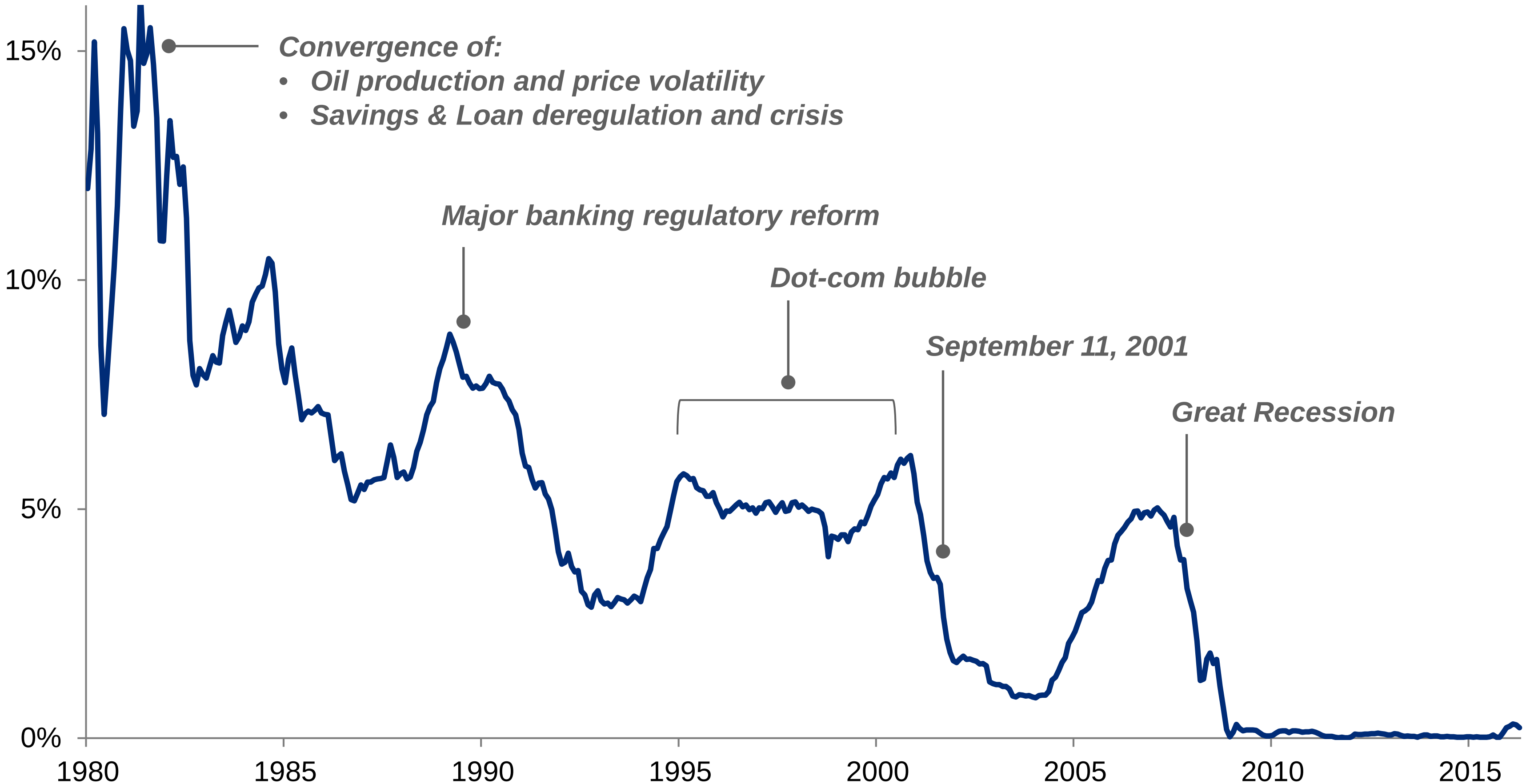
A few pitfalls and how to avoid them

1. Simulating interest rate stress scenarios
2. Applying credibility
3. Understanding the impacts of mix of business

1 Simulating interest rate stress scenarios

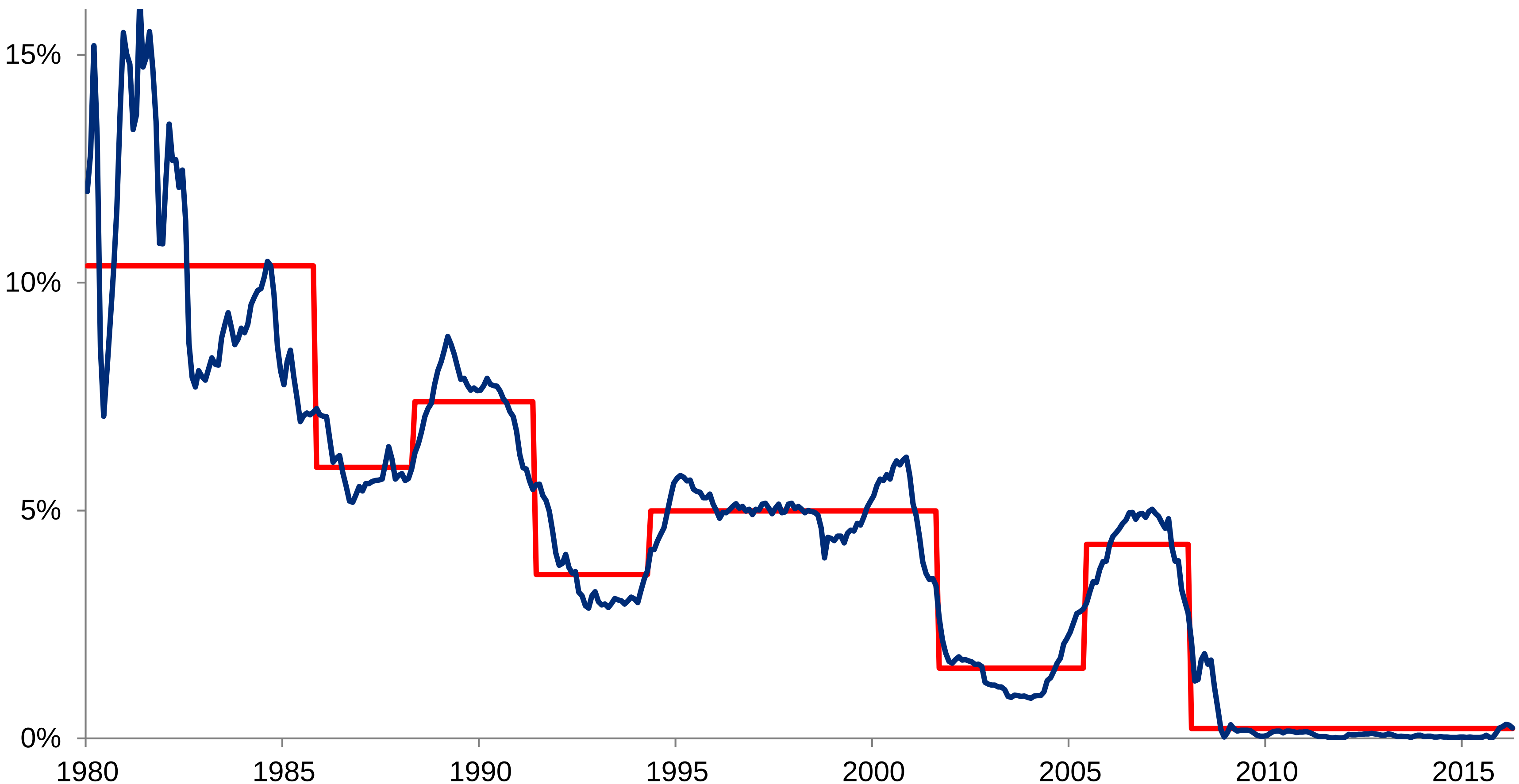
Historical interest rates are not random

Trigger events drive interest rates to new regimes



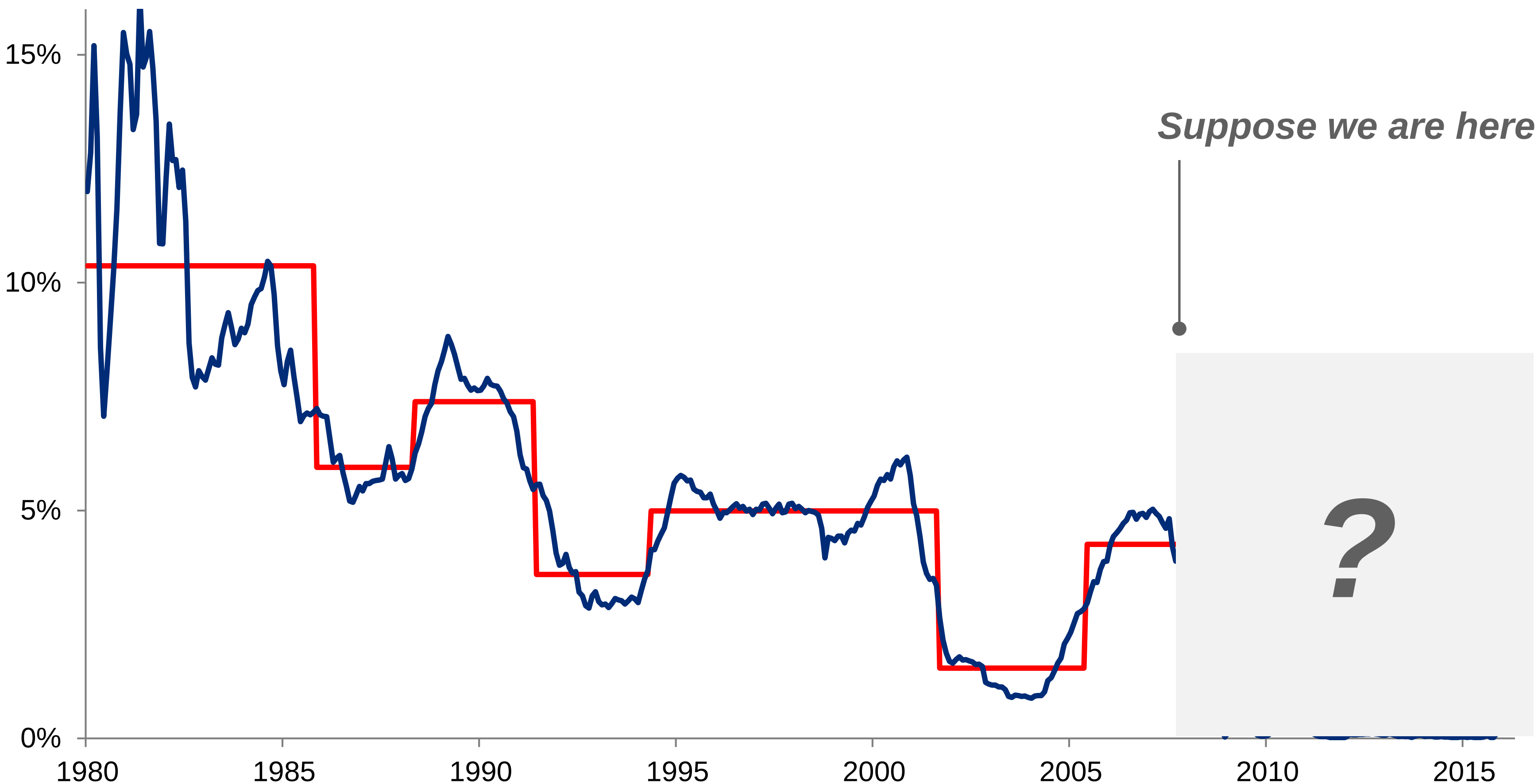
Source: 3-Month Treasury Bill: Secondary Market Rate

Interest rate regimes capture changes to a “new normal”



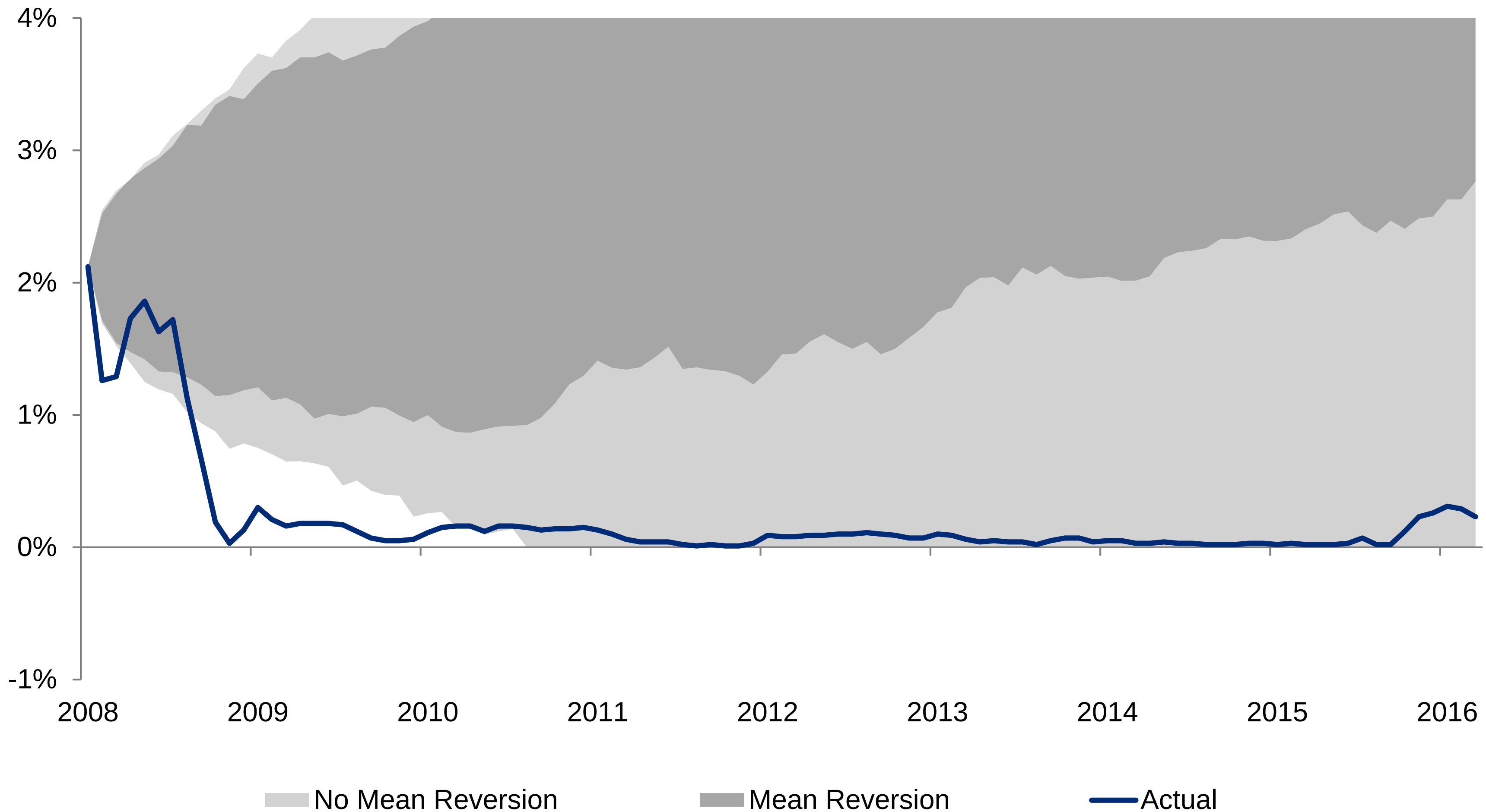
Source: 3-Month Treasury Bill: Secondary Market Rate

Stress scenarios are not predictions
They inform what happens under a range of possibilities



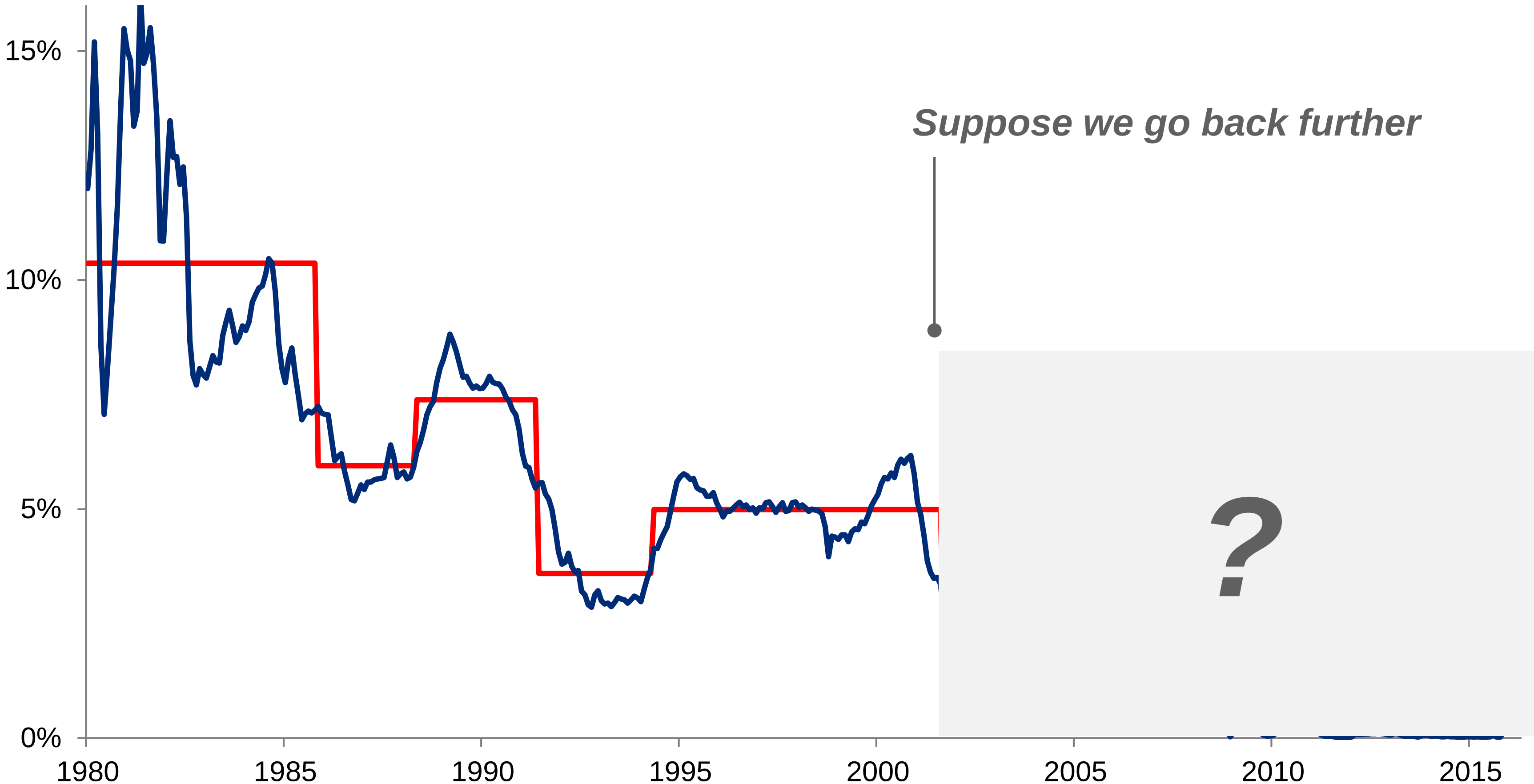
Source: 3-Month Treasury Bill: Secondary Market Rate

Historically calibrated simulations produce a cone of uncertainty
Actual interest rate movements appear more volatile than expected



Source: 3-Month Treasury Bill: Secondary Market Rate

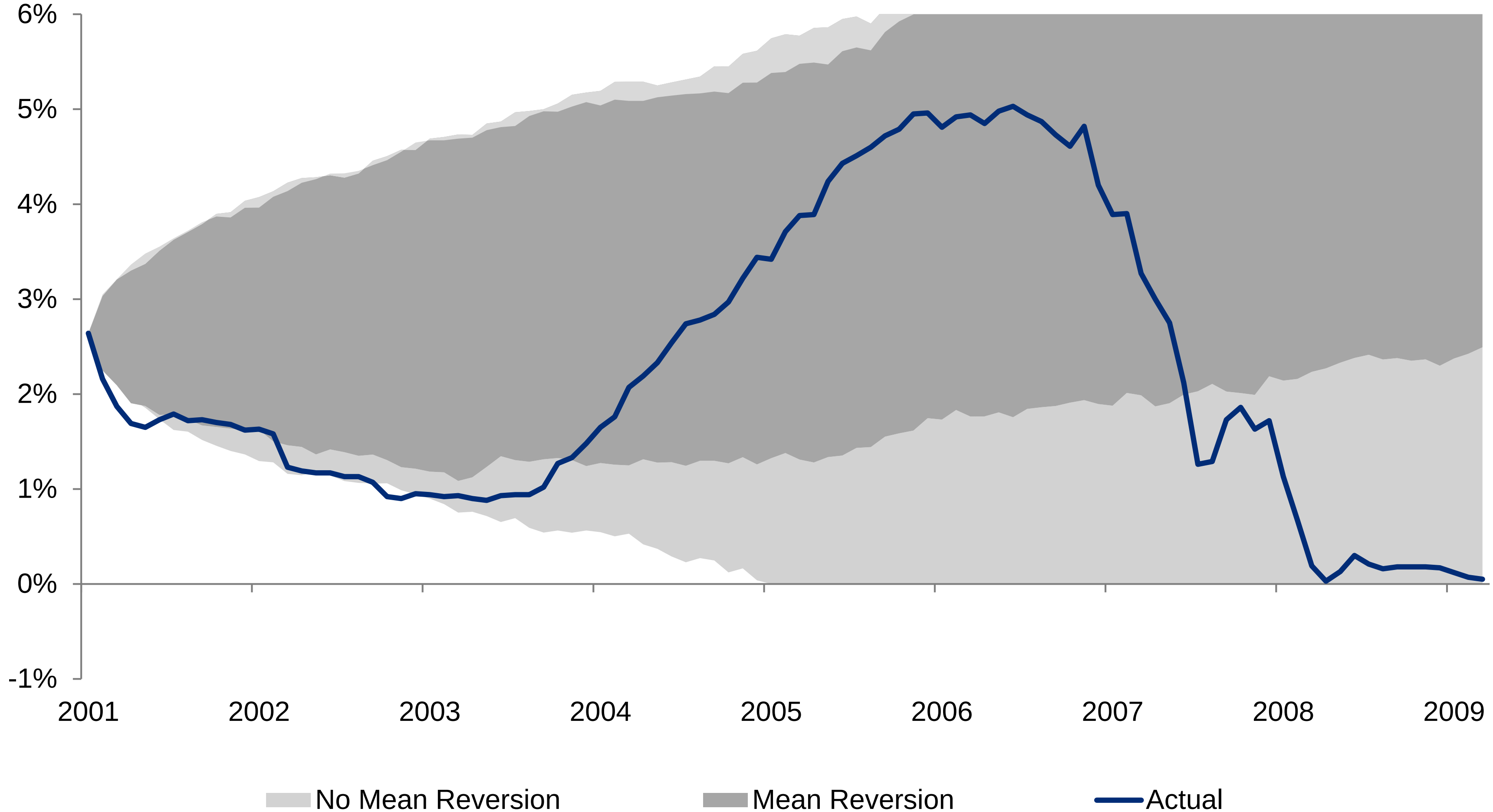
Stress scenarios are not predictions
They inform what happens under a range of possibilities



Source: 3-Month Treasury Bill: Secondary Market Rate

Actual interest rate movement still appears extreme

The simulations are missing the regime-switching feature which leads to underestimating volatility



Source: 3-Month Treasury Bill: Secondary Market Rate

2 | Applying credibility

Credibility parameters are subjectively chosen
The choice of confidence level (p) and margin of error (r) are prone to anchor bias and not selected with a theoretical foundation

Number of deaths needed for full credibility based on r and p (counts)

	r = 1%	r = 3%	r = 5%
p = 90% (z = 1.65)	27,060	3,007	1,082
p = 95% (z = 1.96)	38,416	4,268	1,573
p = 99% (z = 2.58)	66,306	7,367	2,652

Credibility for 1,082 deaths

	r = 1%	r = 3%	r = 5%
p = 90% (z = 1.645)	20%	60%	100%
p = 95% (z = 1.96)	17%	50%	83%
p = 99% (z = 2.575)	13%	38%	64%

Source: SOA Credibility Education Resource for Pension Actuaries August 2017

Credibility is affected by subjective grouping
Subgroups may be partially credible while the total may be fully credible

Mortality study illustration

	Actual	Expected	A/E	Credibility
Male	700	665	105%	80%
Female	400	436	92%	61%
Total	1,100	1,101	100%	100%

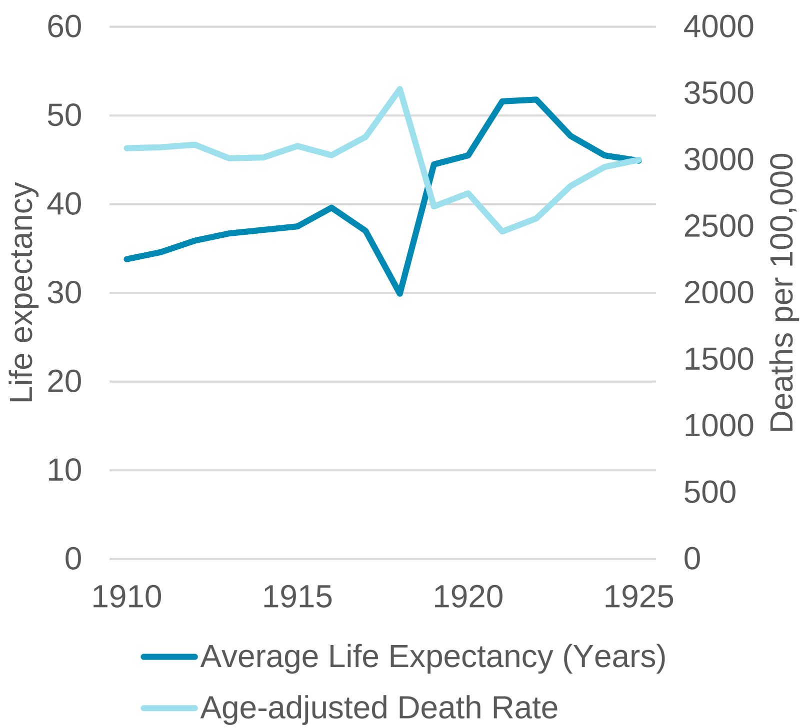
"The whole dataset is fully credible; therefore, no adjustments are needed!"

"The total doesn't matter. The subsets are only partially credible!"

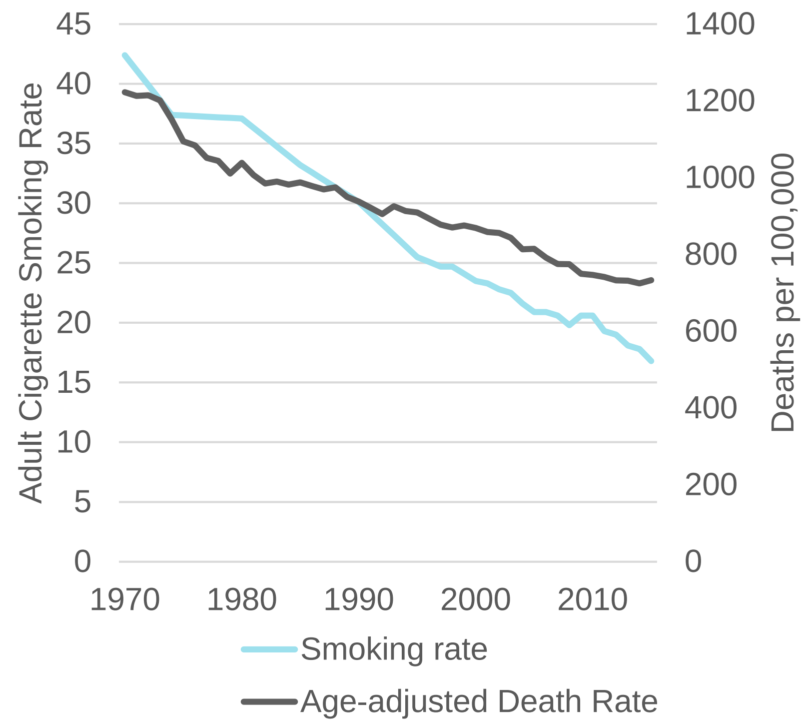
Credibility is not synonymous with predictability

Credibility is a statistical measure that cannot identify drivers or assess to what extent they will continue

1918 Influenza epidemic[1]



Linkage between smoking and deaths[1,2]



First focus on likely drivers and their potential future impacts and only use credibility as a secondary measure of statistical fluctuations.

1. CDC Mortality Trends in the United States, 1900-2015
2. American Lung Association, Adult Cigarette Smoking Rate Trend by Sex, Race, and Age

3 Understanding the impacts of mix of business

High-level results may be misleading (1 of 4)
Equal counts for male and female data

Illustration of Simpson's paradox

	Male		Female	
	Count	IRR	Count	IRR
Total	1,200	5%	1,200	4%

Naïve conclusion #1: On average, males have greater IRR than females

High-level results may be misleading (2 of 4)
Changing mix of business over time actually reverses previous conclusion

Illustration of Simpson's paradox

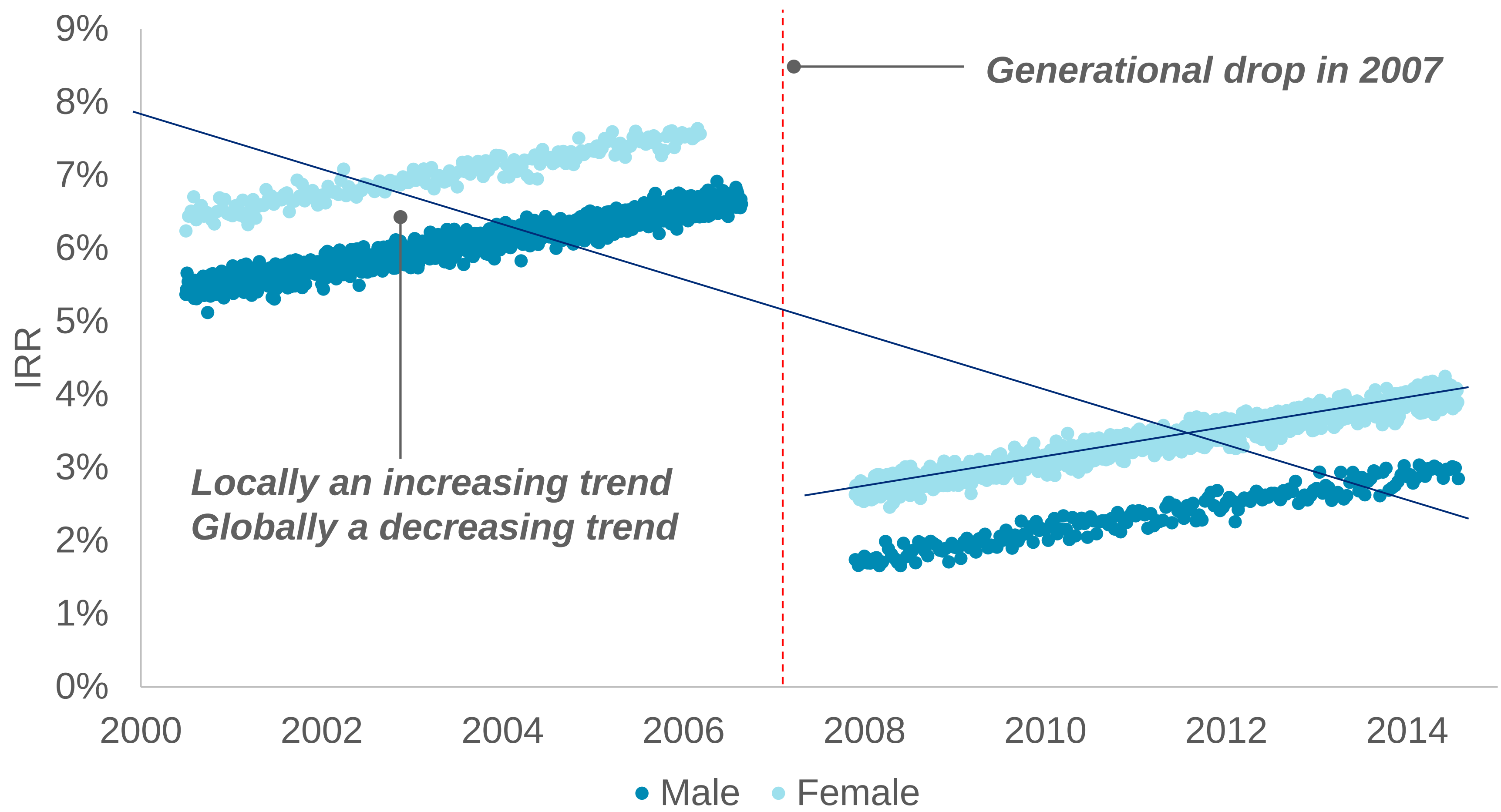
	Male		Female	
	Count	IRR	Count	IRR
Pre-2007	1000	6%	200	7%
2007+	200	2%	1,000	3%
Total	1,200	5%	1,200	4%

Naïve conclusion #2: For both genders, the IRR is decreasing over time

High-level results may be misleading (3 of 4)

Graphical representation shows clear linear trends in each cohort, reversing previous conclusion

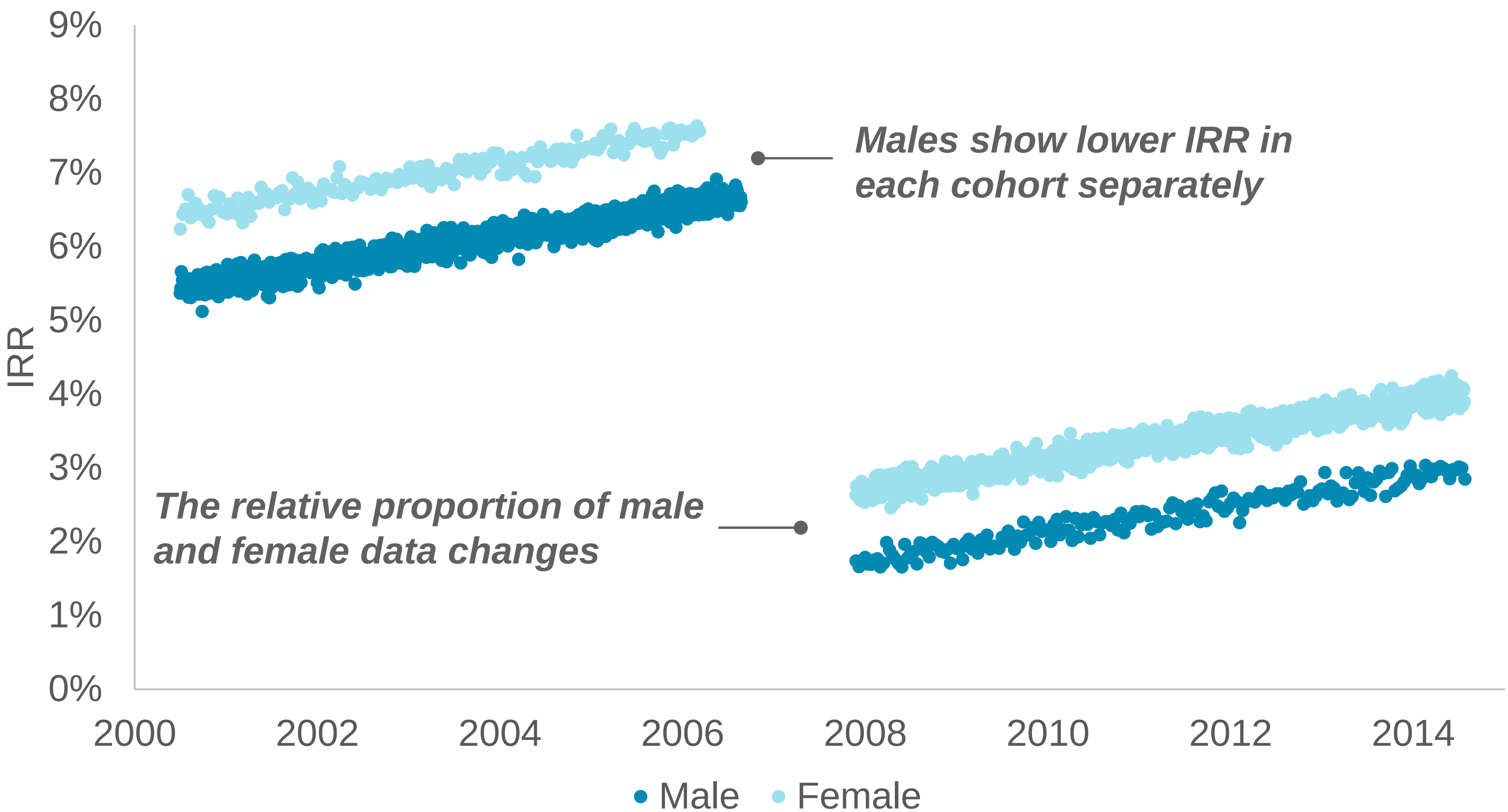
Illustration of Simpson's paradox



High-level results may be misleading (4 of 4)

Generational gaps and changing mix of business can produce counterintuitive results

Illustration of Simpson's paradox



Lessons learned

Traditional interest rate models underestimate volatility

- Historical interest rates are not random walks, but are driven by real-world factors that act as triggers
- Calibrating purely to historical rates ignores these real-world regime changes
- For stress testing, or risk appetite analysis, regime switching interest rate models may capture volatility better than non-regime switching models

Credibility is (much) less objective than it appears to be

- Credibility is a formulaic framework that may give users a false sense of security
- Users may be making decisions based on somewhat hidden subjective decisions without being fully aware (or disclosing) the implications
- Full credibility does not mean full predictability

A change in the mix of business may produce counterintuitive results

- Trends in aggregate may not be indicative of trends within subgroups
- Generational shifts need to be well understood
- Data visualization can reveal patterns about trends, data density and help prevent naïve or false conclusions