LEARNING TO BE OVERPRECISE

CHRISTOPH MERKLE

LEIBNIZ OPEN SCIENCE DAY 2024
25TH NOVEMBER 2024

DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

12 SEPTEMBER 2023

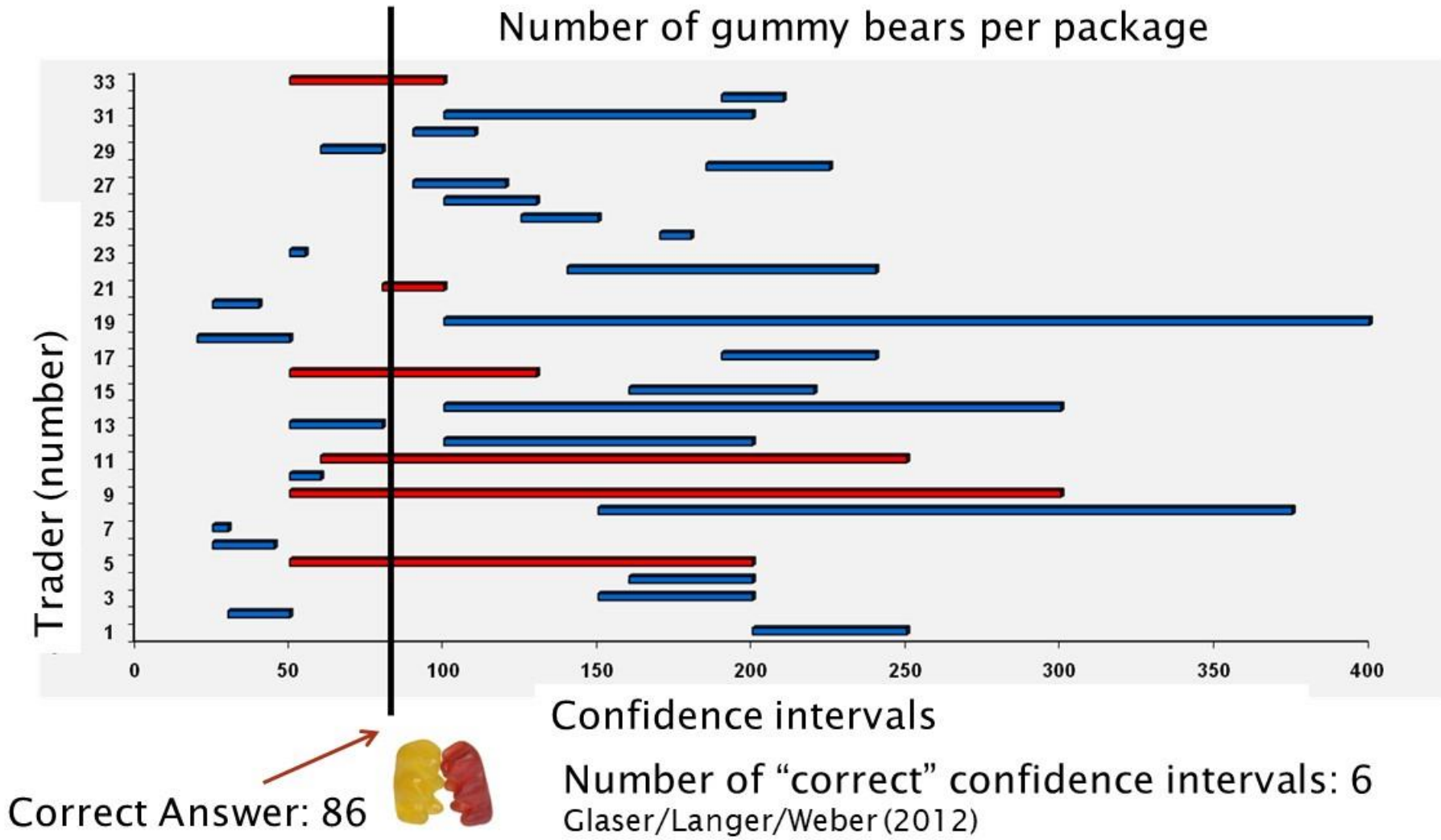
CHRISTOPH MERKLE
ASSOCIATE PROFESSOR

HOW MANY GUMMY BEARS?

In what range do you believe the true value falls with a probability of 90%?

Between ____ and ____

RESPONSE BY FINANCIAL EXPERTS



OVERCONFIDENCE

THREE (MAIN) TYPES OF OVERCONFIDENCE

1 OVERESTIMATION

People can be overconfident with regard to their **absolute ability** or performance in a domain

EXAMPLE

People underestimate the time they will need to run a Marathon (Grieco and Hogarth 2009)

2 OVERPLACEMENT

People can be overconfident with regard to their **relative ability** or performance in a domain compared to others

EXAMPLE

~90% of American drivers rate themselves as more skilled than the median driver (Svenson 1981)

3 OVERPRECISION

FOCUS

People overestimate the **precision of their knowledge** and forecasts

EXAMPLE

Gummy bears

AIM OF THE TALK

1) A perspective on individual replication studies in view of the rise in large scale replication studies (many labs studies), crowd-sourced studies, megastudies, or metastudies

   "...a single design is largely uninformative about whether or not the underlying hypothesis is supported." Huber et al. (2023)

2) Methodology and results of our paper (in brief)

   Merkle, C., Schreiber, P. (2024): Learning to be overprecise. *Journal of Business Economics*, accepted. https://doi.org/10.1007/s11573-024-01203-w

WHICH PAPERS TO REPLICATE?

A common issue with replication studies:

- Mostly high-impact articles are replicated (Mueller-Langer et al., 2019)
- *"To contribute data about replicability in economics, we replicated 18 studies published in the American Economic Review and the Quarterly Journal of Economics between 2011 and 2014."* Camerer et al. (2016)
- *„Take the set of papers that are about 10-15 years old, that have not yet been shown to have serious issues, and that have gathered a lot of citations in the last 24 months"* (guidelines Critical Finance Review)

➔ Does it justify costs and effort to replicate "low impact" studies?

Deaves, R., Lüders, E., & Schröder, M. (2010). The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior & Organization*, *75*(3), 402-412.

305 cites (Google scholar), 134 cites since 2020

Boutros, M., Ben-David, I., Graham, J. R., Harvey, C. R., & Payne, J. W. (2020). The persistence of miscalibration. *National Bureau of Economic Research*, Working Paper No. w28010.

30 cites (Google scholar)

A RISK WORTH TAKING?

Replications may not publish well (in particular, if they are successful)

- Example: The infamous Reinhart and Rogoff (AER, 2010) replicated by Herndon, T., Ash, M., & Pollin, R. (Cambridge Journal of Economics, 2014)
- Creates “overturn bias” (Galiani, Gertler, and Romero, 2017)
- Little self-correction of science after replication (von Hippel, 2022)

In our case, a special issue explicitly called for replication studies in finance:

PRE-REGISTERED REPLICATION

An ideal (?) process:

The process of the SI in practice:

- Submission of extended abstract (deadline 31/5/2023)
- Feedback on extended abstract with invitation for full article (23/6/2023)
- Draft pre-registration based on editorial feedback (7/8/2023)
- Editor declined comment on pre-registration, non-committal (8/8/2023)
- Submission of article (30/11/23)
- Reviewer feedback and R&R (05/03/2024)

PURPOSE OF REPLICATION

Replications can help to advance theory, advance methodology, establish robustness/generalizability, document replication failures...

1) Theory
   - Rational (Bayesian) updating of beliefs or self-attribution bias
   - Misinterpretation of critical result (DLS)

2) Methodology
   - Problematic extraction of one-months forecasts from six-month forecasts, invalid assumptions (DLS)
   - Issues with standard errors (DLS), fixed effects (BBGHP)

3) Data / Robustness
   - Short time-series (DLS)
   - High fluctuation among participants (BBGHP)

DATA

ZEW Financial Market Survey:

> 6b. Six month ahead, I expect the DAX to stand at [_____] points. With a probability of 90 per cent the DAX will then range between [_____] and [_____] points.

- Monthly survey, panel data from 2003 – 2022
- Financial professionals from banks, insurance companies and corporations
- About 200 participants in each survey wave, 785 unique participants

In addition: Duke CFO survey, 80% CIs for S&P 500, quarterly data, 2001-2018

Data in DLS: two year of data from the ZEW survey

➔ extension, out of sample analysis, update

UPDATING BELIEFS

How do we expect participants to update their beliefs?
Two general directions:

Rational Updating

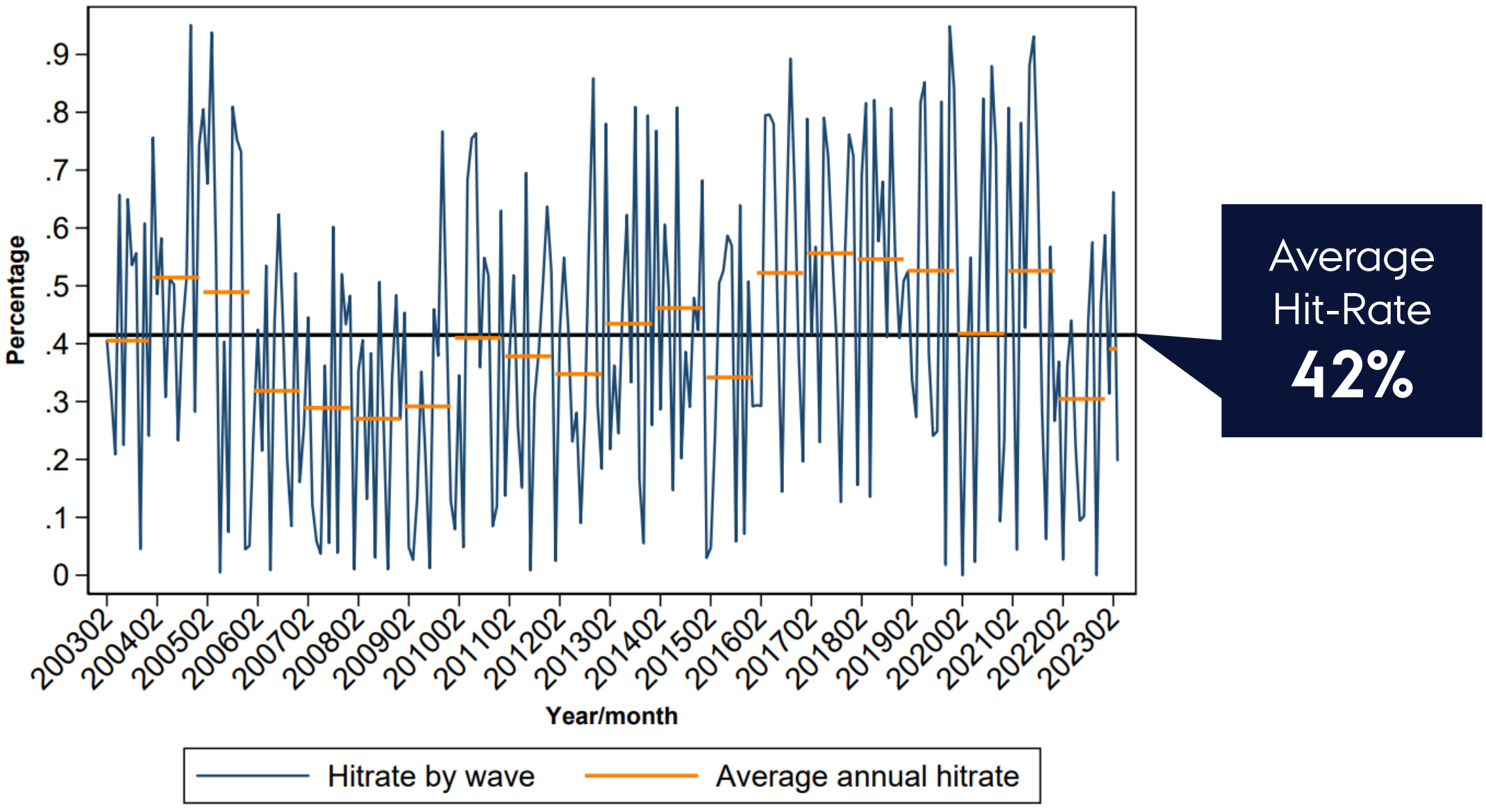
- Bayesian updating (BBGHP)
- Confidence interval reflects prior conviction
- Hit / miss as new information (signal) lead to updated beliefs
- A hit is a weaker signal than a miss (should occur 90% of the time)
- Bayesian updating can lead to proper calibration over time

Biased Updating

- Self attribution bias as potential explanation for overprecision
  (e.g., Kahnemann 2011, Moore et al. 2016)
- A person attributes positive outcomes to their own skills, but blames factors outside their control for negative outcomes
- Hits are attributed to own skill, misses to bad luck
- Learning to be overconfident

OVERCONFIDENCE OVER TIME

Financial professionals as a group are overprecise



RESPONSE TO HITS AND MISSES

Learning takes place in the sense that **after hits confidence intervals contract and after misses confidence intervals expand**

	Δ CI Width			
	(1)	(2)	(3)	(4)
$Miss_{t-6}$	3.37***	3.40***	3.78***	3.14***
	(0.16)	(0.16)	(0.18)	(0.15)
Unexpected Vol.				0.30***
				(0.01)
Exp. Change in Vol.				0.35***
				(0.05)
Constant	−2.88***	−2.49***	−17.55***	−2.12***
	(0.16)	(0.10)	(1.06)	(0.09)
R^2	0.02	0.02	0.23	0.13
Observations	35775	35775	35775	35775
Forecaster Fixed Effects	N	Y	Y	Y
Time Fixed Effects	N	N	Y	N

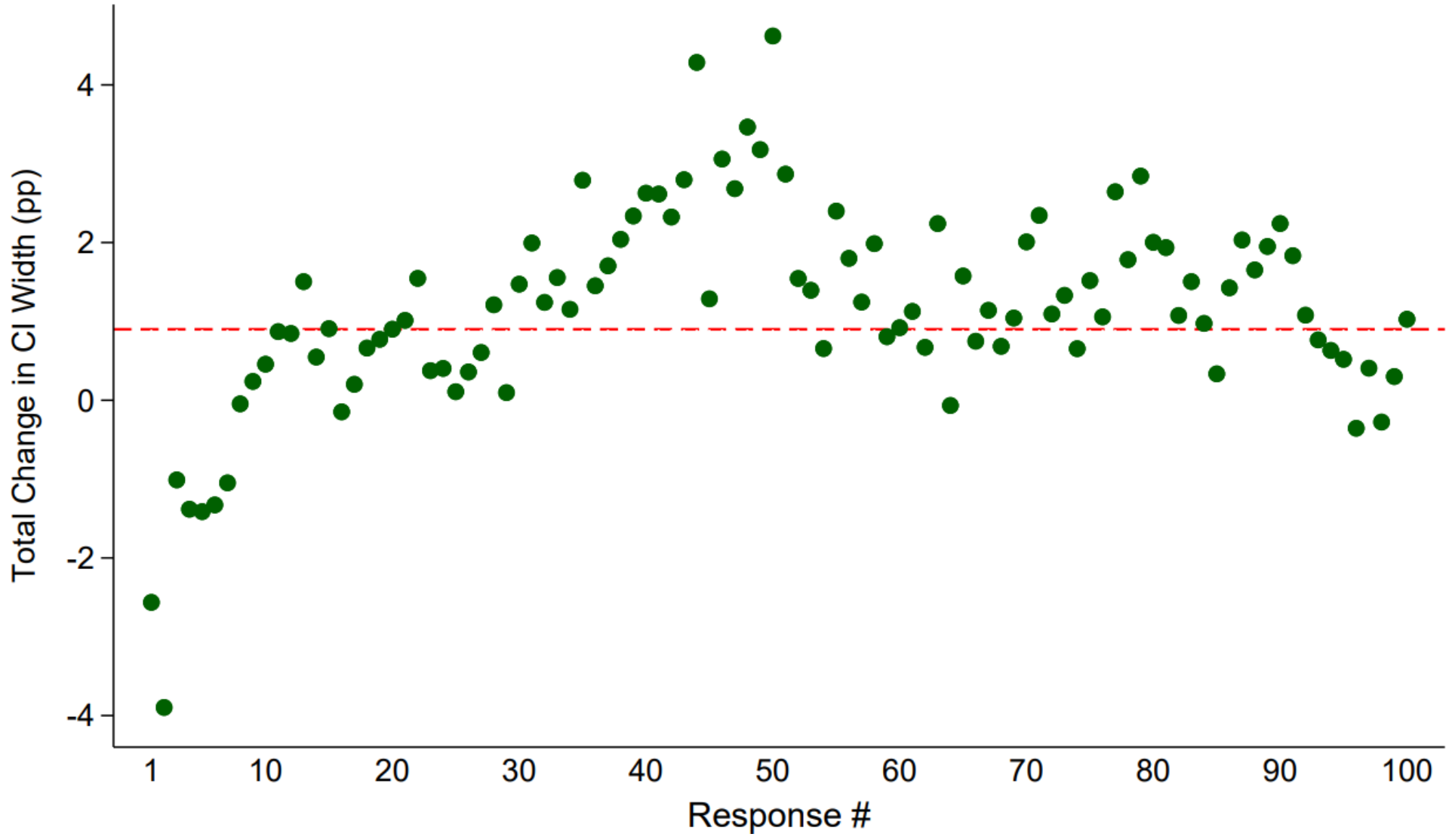
HOW TO DISTINGUISH EXPLANATIONS

Rational Updating		Biased Updating
• The reaction to a miss is stronger than the reaction to a hit since the signal is more informative	⟷	• The reaction to a miss is not necessarily stronger, as misses are attributed to bad luck
• Subsequent misses lead to **less adjustments** in CIs as weight on prior increases	⟷	• Accumulating feedback can lead to finally realizing that narrow CIs are untenable
• Even narrow misses of confidence intervals should typically give rise to increasing CIs	⟷	• Narrow misses might be classi-fied as “almost” correct or other forms of self-serving hindsight
• Misses on the upside challenge the confidence interval just as much as misses on the downside	⟷	• Misses on the upside occur in exuberant markets which exacerbate overconfidence

HITS AND MISSES

Panel A: ZEW sample	Regression specifications from Table 5			
	(1)	(2)	(3)	(4)
Total Δ CI width	− 2.88	− 2.49	− 2.48	− 2.46
	(0.16)	(0.10)	(0.61)	(0.13)
Total %Δ CI width	− 12.6%	− 10.9%	− 10.9%	− 10.8%
Panel B: Duke CFO sample	Regression specifications from BBGHP, Table 4			
	(1)	(2)	(3)	(4)
Total Δ CI width	− 3.07	− 3.74	− 4.09	− 3.70
	(0.32)	(0.37)	(1.29)	(0.43)
Total %Δ CI width	− 13.4%	− 16.4%	− 17.9%	− 16.2%
For comparison change after miss				
Total Δ CI width (original)	1.61	1.88	1.99	1.75
Total Δ CI width (replicated)	1.56	1.84	1.94	1.85
Total %Δ CI width (original)	13.0%	15.2%	16.2%	14.2%
Total %Δ CI width (replicated)	12.9%	15.1%	16.0%	15.2%

ADJUSTMENT IN THE LONG RUN



NARROW MISSES

Total Δ CI Width	(1)	(2)	(3)	(4)
Near miss (<1pp)	–0.77	–0.34	–0.31	–0.33
Miss by <1pp	–0.66	–0.23	–0.17	–0.22
Miss by 1pp to <2pp	–0.98	–0.58	–0.51	–0.55
Miss by 2pp to <3pp	–0.87	–0.45	–0.42	–0.44
Miss by 3pp to <4pp	–0.60	–0.16	–0.09	–0.15
Miss by 4pp to <5pp	–0.29	0.12	0.19	0.14
Miss by 5pp to <6pp	–0.27	0.15	0.24	0.18
Miss by 6pp to <7pp	–0.08	0.34	0.39	0.34
Miss by 7pp to <8pp	0.20	0.64	0.62	0.61
Miss by 8pp to <9pp	–0.16	0.29	0.31	0.27
Miss by 9pp to <10pp	0.91	1.35	1.45	1.35
Miss by >10pp	1.84	2.30	2.30	2.23

UPSIDE VS. DOWNSIDE

	Δ UCI			Δ LCI		
	(1)	(2)	(3)	(4)	(5)	(6)
Miss High						
Total Δ CI Width	–0.60	–0.45	–0.47	–0.70	–0.51	–0.52
	(0.06)	(0.37)	(0.07)	(0.09)	(0.49)	(0.09)
Total %Δ CI Width	–10.3%	–7.7%	–7.9%	–8.2%	–5.9%	–6.1%
Miss Low						
Total Δ CI Width	2.04	2.13	2.16	2.90	3.06	3.07
	(0.12)	(0.44)	(0.12)	(0.18)	(0.57)	(0.17)
Total %Δ CI Width	43.6%	45.5%	45.9%	41.6%	44.0%	44.0%
Forecaster Fixed Effects	N	Y	Y	N	Y	Y
Time Fixed Effects	N	Y	N	N	Y	N

CONCLUSION

We replicate two papers on the dynamics of overconfidence

- We can confirm high overprecision and no learning on the aggregate for financial professionals when predicting the stock market
- We further confirm updating in the correct direction, but insufficient to reach proper calibration
- This is mainly due to an equally strong reaction to hits and misses, which has participants bouncing around 50% rather than ever reaching 90%
- This and other findings is best explained by biased self-attribution

One of the papers we replicate was a working paper

➔ We shared our results with the authors

➔ The paper is now R&R at the RFS, new version as of 6 Nov 2024

➔ Does not acknowledge or cite our work

DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY