Experience Credibility from Account Characteristics

A Logistic Extension of Bühlmann-Straub with Temporal Adaptation

Jake Morris FIA CSPA · Head of Global Pricing & Portfolio Analytics, Allianz Commercial
Insurance Data Science Conference 2026

"Three outputs. One likelihood."

Bühlmann-Straub: powerful but flawed in some settings

Bühlmann-Straub[†] Default Formulation

$$\hat{r}_i = (1 - Z_i)\mu + Z_i \bar{f}_i$$

$$Z_i = \frac{E_i}{E_i + K}$$

Explicit credibility weight ✓
Easy to communicate ✓
Simple to implement ✓

Key Weaknesses

1. Single pooled "K"[††] ✗
2. All years carry equal weight ✗
3. Flat complement ✗
4. *[No measure of uncertainty]*

What about GLMMs?

- Handle heterogeneity (flexible random effects) ✓
- Complement estimated jointly ✓
- Difficult to implement & communicate ✗

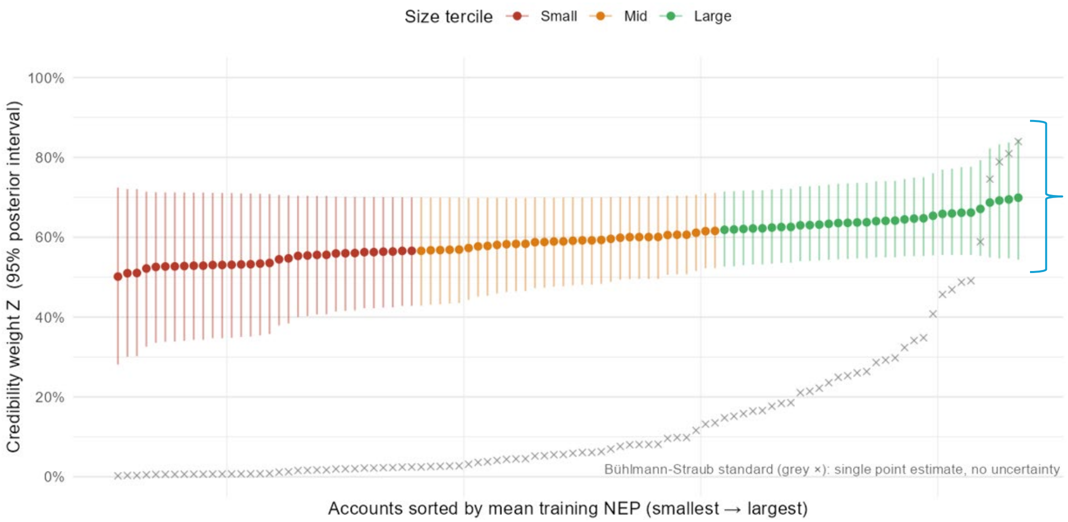
Requires accounts in training history: new commercial submissions out of reach

Practitioner Patches

1. Stratify or manually set K
2. Manually downweigh old experience[†††]
3. Provide an external complement

No Unified Framework

Proposed Framework Resolves Weaknesses Simultaneously



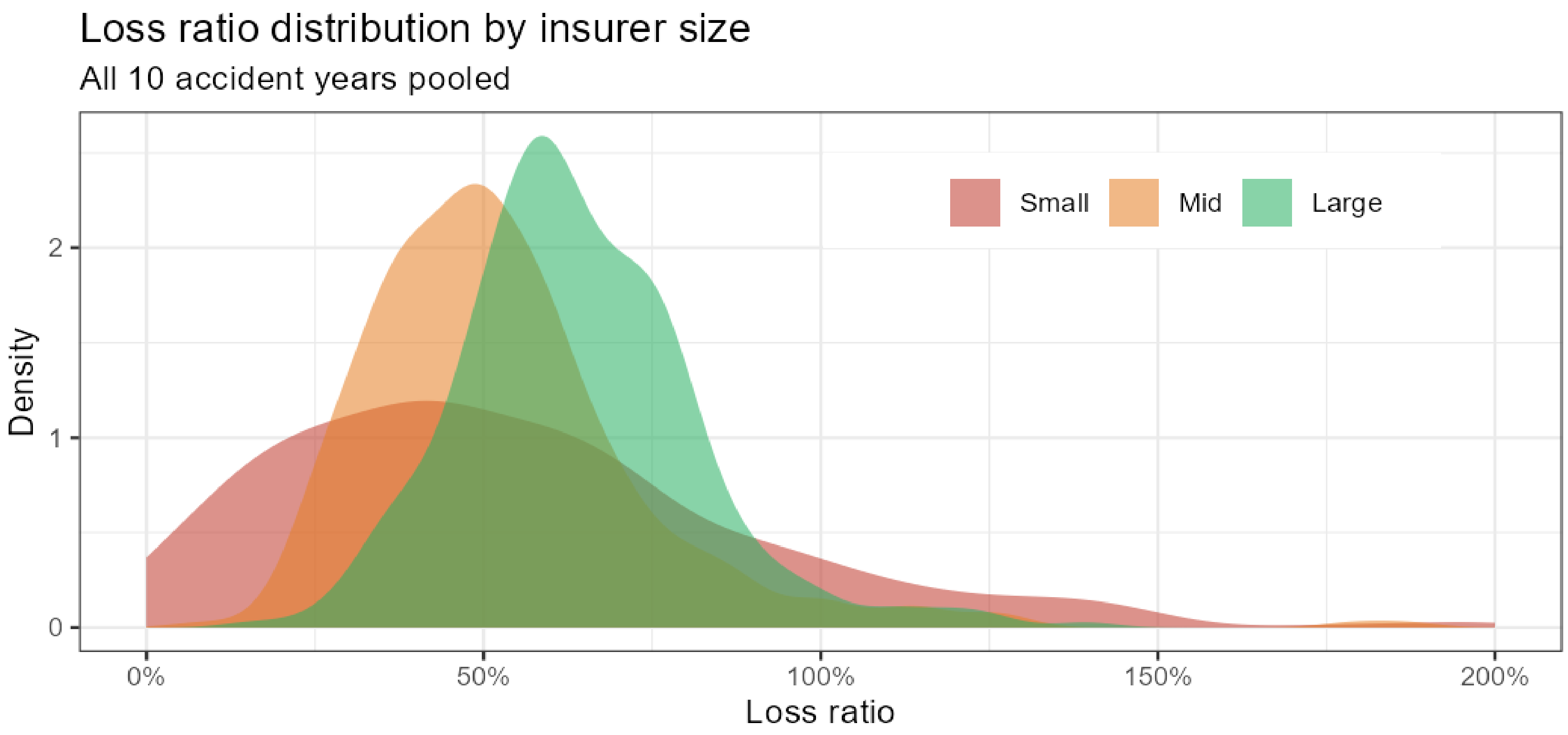
Bonus: Optional Z uncertainty quantification (UQ)

[†]B-S [††]$K = \sigma_\varepsilon^2 / \sigma_u^2$ = ratio of within to between account variance

[†††]E.g. Exponentially Weighted Moving Average (EWMA)

Commercial data findings: K, μ, λ† vary by account size

CAS Commercial Auto · 96 Companies · AY 1998- 2007 Train AY 2001-05 | Test: AY 2006-07 | Response: company LR ÷ portfolio LR

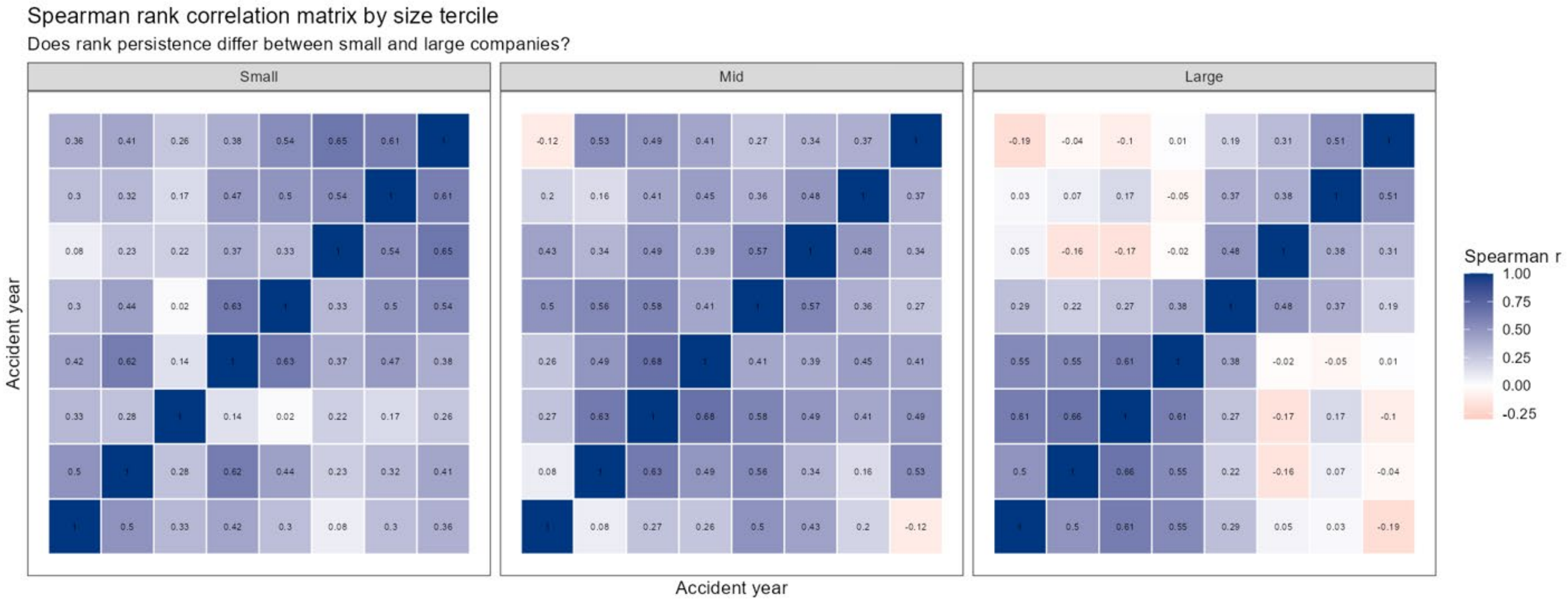


Average LR†† varies by size | Optimal "K" varies by size

†Experience temporal discounting (defined on slide 5) ††Loss Ratio

Commercial data findings: K, μ, λ† vary by account size

CAS Commercial Auto · 96 Companies · AY 1998- 2007   Train AY 2001-05 | Test: AY 2006-07 | Response: company LR ÷ portfolio LR



LR correlation decay faster for large → optimal lookback varies by size

†Experience temporal discounting (defined on slide 5)

One likelihood: nesting, flexibility, and joint estimation

1. Credibility Weight Sub-Model

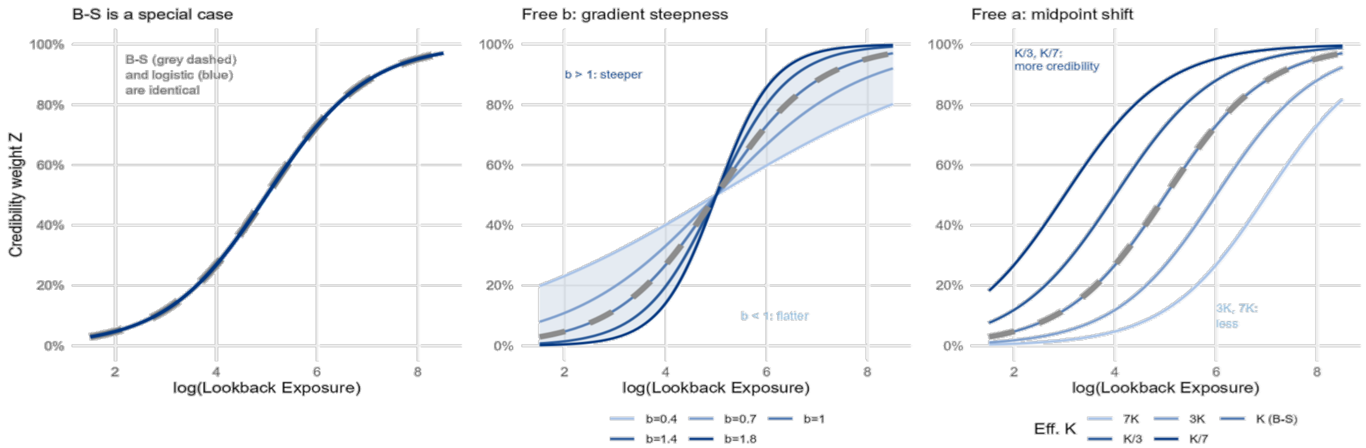
$$Z_i = logistic\left(a_i + b_Z log(\widetilde{E}_i)\right)$$

$\widetilde{E}_i = \sum_{k=1}^{W} E_{i,t-k}$

$E_{i,t-k}$ is "lookback exposure" for account i at time t-k

a_i can include other covariates
e.g. industry, tenure, volatility, …

When a_i is a single intercept, a_Z:
$K_i = e^{-a_Z} \widetilde{E}_i^{1-b_Z}$



2. Temporal Decay Sub-Model

$$\hat{f}_i^{(\lambda)} = \frac{\sum_k \lambda^{k-1} C_{i,t-k}}{\sum_k \lambda^{k-1} E_{i,t-k}}$$

$C_{i,t-k}$ is the loss statistic of interest for account i at time t-k

$\lambda \in [0, 1]$ determines experience temporal decay†

3. Model Components Estimated Simultaneously (optionally including complement)

$$\hat{r}_i = (1 - Z_i)\mu_i + Z_i \hat{f}_i^{(\lambda)}$$

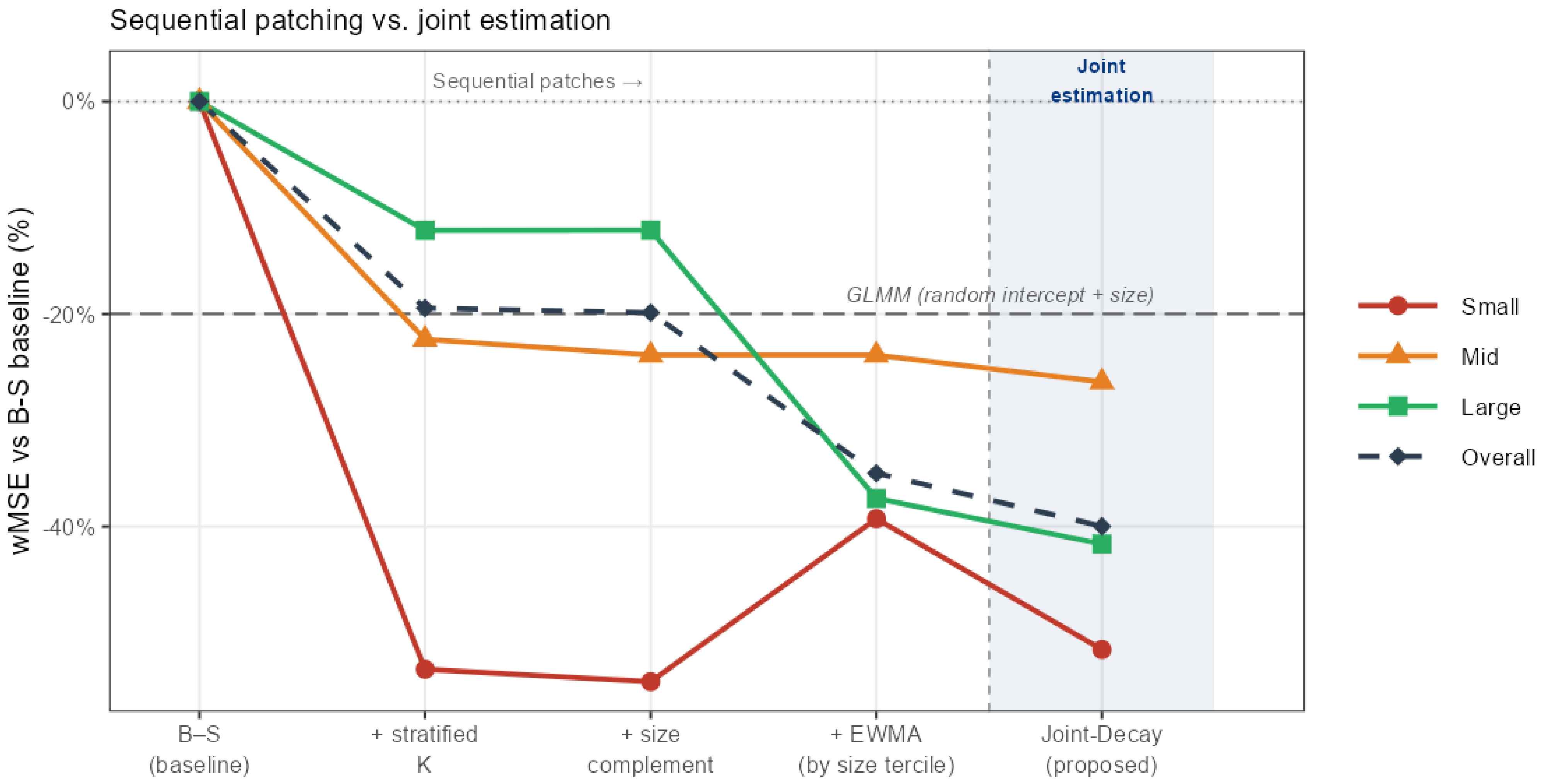
Requires a likelihood — B-S is distribution-free. Gamma (LR/sev), Poisson/NB (freq), Tweedie (LR/PP). Enables joint estimation and UQ; must be validated.

Set $\boldsymbol{b_Z}=1$, $\boldsymbol{a_i}=-\ln K$, $\boldsymbol{\lambda}=1$, $\boldsymbol{\mu_i}=\mu$ → Bühlmann-Straub exactly††

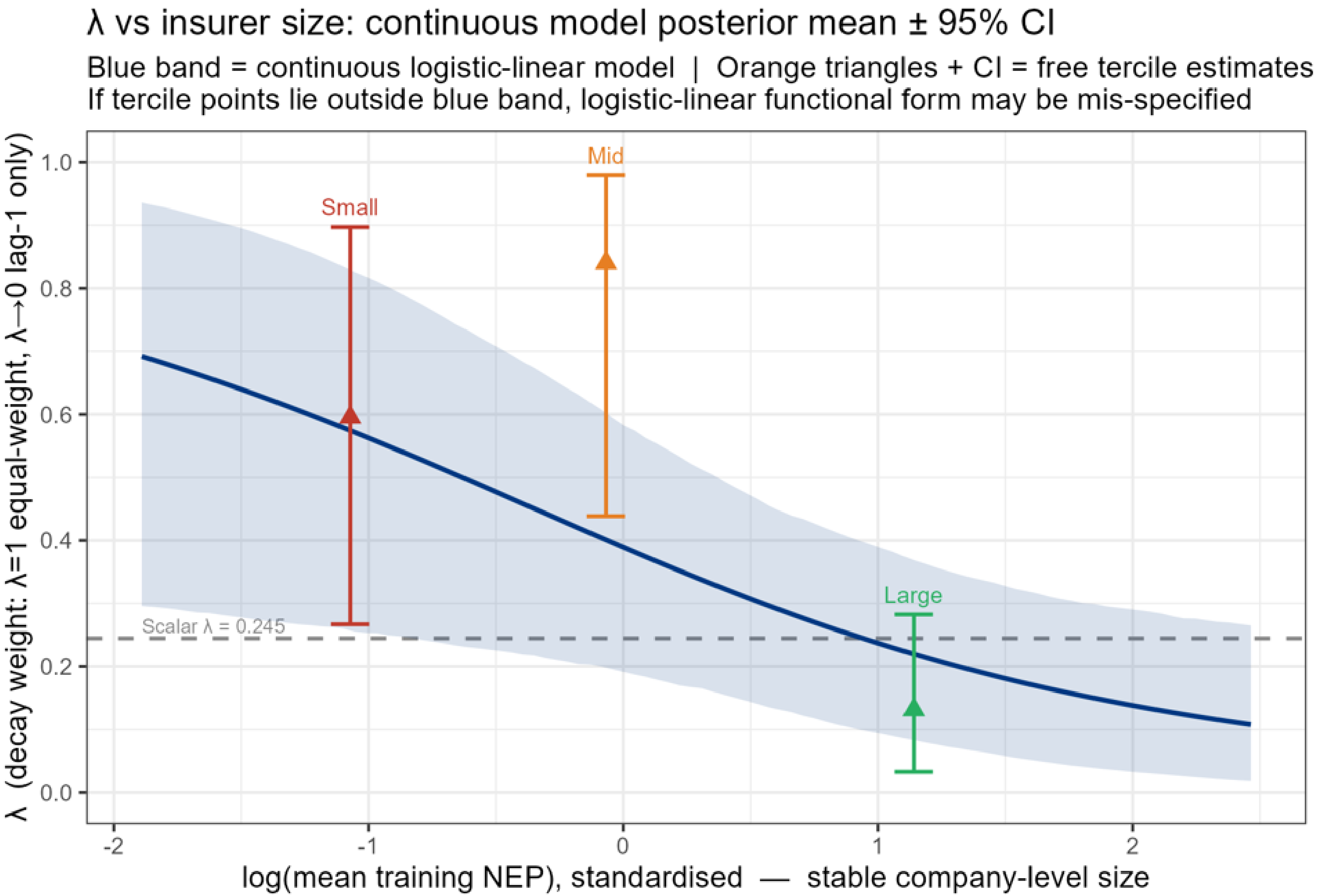
†λ_i an also be estimated from account characteristics

††using lookback exposure

Joint Model Results: best held-out wMSE of 20 variants

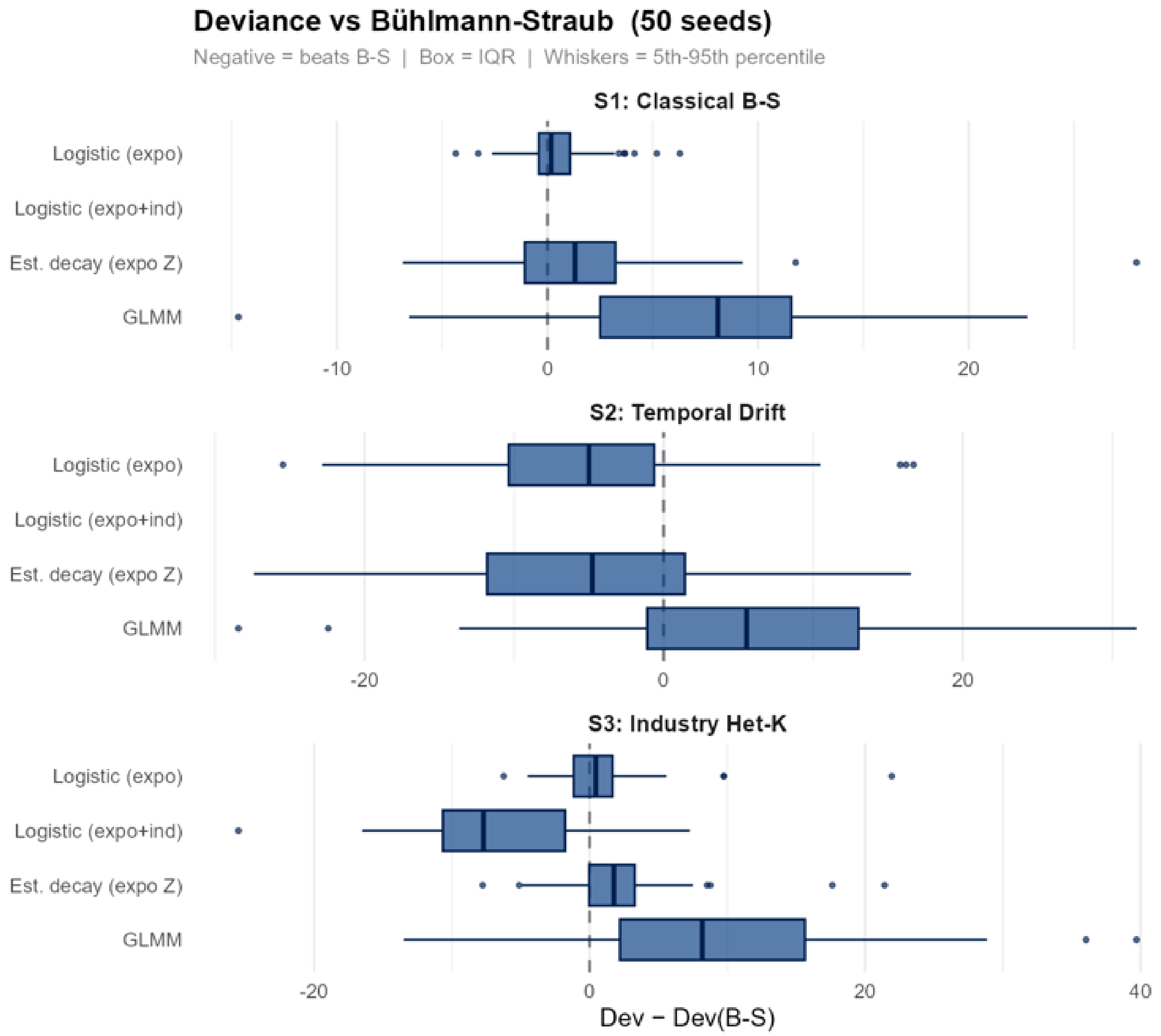


Optimal experience lookback depends on account size



Large: λ≈0.13 (~1yr) | **Small: λ≈0.60 (~2.5yr)** | **Mid: λ≈0.84 (~6yr)** *Scalar λ (dashed) misses the gradient entirely · Replicates qualitatively on Other Liability*

Simulation: finds signal where exists – quiet if it doesn't



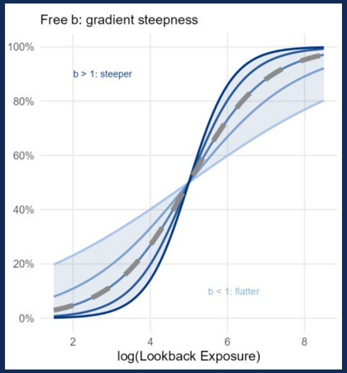
- **S1 stable risk:** neutral, no spurious gain
- **S2 risk drifts:** Joint-Decay 1st, 82% of seeds
- **S3 industry drives Z:** recovers only when included

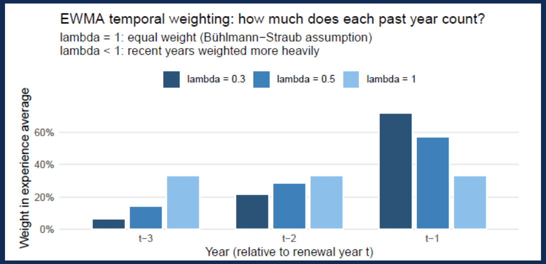
x-axis: deviance vs B-S · negative = beats B-S

Logistic Credibility with Temporal Decay Summary

Three Structural Properties

1. **Nests Bühlmann-Straub exactly**† – gain from flexibility, not method switch
2. **Temporal adaptation estimated** from data – no pre-assumed lookback
3. **Posterior uncertainty over Z** (not shown) – identify reasonable ranges for credibility weights





This Case Study

- **Best wMSE** of 20 models on held-out test
- **−38%** wMSE vs B-S · **+15pp** Gini
- **Calibration corrected** all size groups
- **UQ††: 89–91%** coverage (vs 67% constant ϕ)

Resources

1. **Pre-print paper:** https://arxiv.org/abs/2606.08692
2. **GitHub repository** with R/Python code: https://github.com/jakem87/logistic-credibility
3. **Contact:** jake.morris87@gmail.com

†For B-S with "lookback exposure"

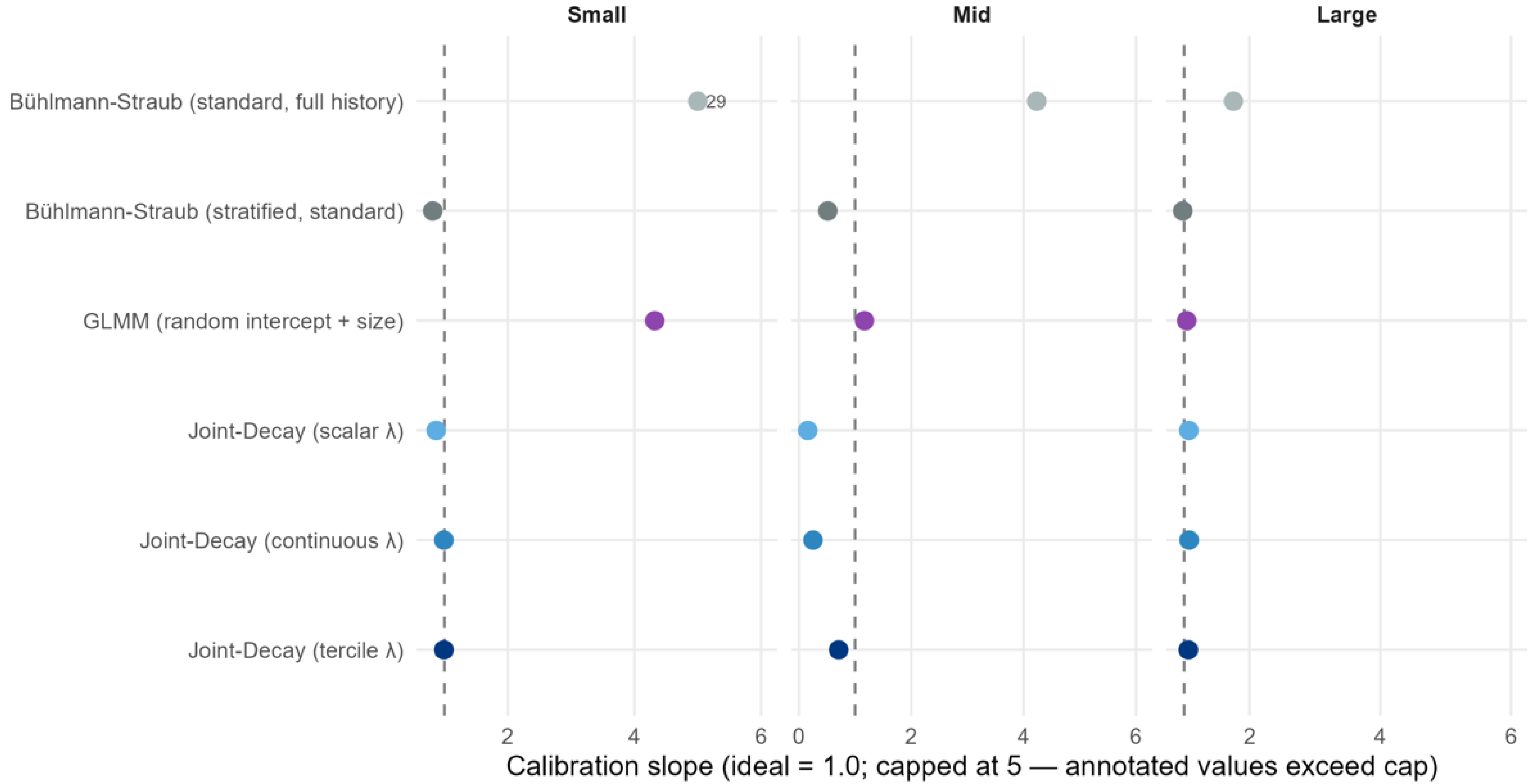
††Uncertainty quantification (optional): posterior predictive intervals

Appendix

Key Case Study Results: 10 model comparison

Model	wMSE ($\times 10^{-3}$)	log-wMSE ($\times 10^{-3}$)	Gini_{pct}	Calib. slope
Baselines				
Market Mean (baseline)	19.25	70.82	0.6%	1.00
Last Year LR (naive)	13.14	49.59	75.1%	0.63
B-S benchmark				
B-S (standard, full history)	13.27	55.38	63.8%	1.81
Best B-S sequential patch				
B-S (strat K + size comp, tercile λ)	8.62	33.77	76.7%	0.89
Best GLMM competitor				
GLMM (random intercept + size)	10.62	45.20	69.7%	1.06
Logistic credibility (comparators)				
Joint-Decay (scalar λ)	8.61	34.98	76.5%	1.00
Joint-Decay (continuous λ)	8.38	34.32	77.2%	1.02
Proposed model				
Joint-Decay (tercile λ)	**7.96**	**32.65**	**78.7%**	**1.03**
Diagnostic: B-S Z variant				
Joint-Decay (B-S Z + tercile λ)	9.00	36.79	79.2%	0.78
Dispersion extension†				
Joint-Decay (tercile λ + disp.)	8.11	33.17	80.1%	1.09

Joint Model Results: calibration slope by size tercile



Held-out predictions: models fitted on AY 2001–2005. Dashed line = perfect calibration (slope = 1.0). B-S Small slope ~29 is truncated.

Uncertainty Quantification: Z vs. Predictive Intervals