



TASK ACTUARIAL CONVENTION

THE NEXT FRONTIER—
ACTUARIES BEYOND
BOUNDARIES



Supervising the Chain Ladder

James Grove, Anesu Shuro
Psicle | Dynamo Analytics

#TASKConvention2026

convention.actuarieskenya.or.ke

Machine learning grew out of statistics



Fitting

OLS



Ridge / lasso

Tabular data

Linear models



Tree ensembles

Time series

ARIMA



RNNs / transformers

Images

Hand-crafted features



Convolutional neural networks

Reserving

Chain ladder / Cape Cod



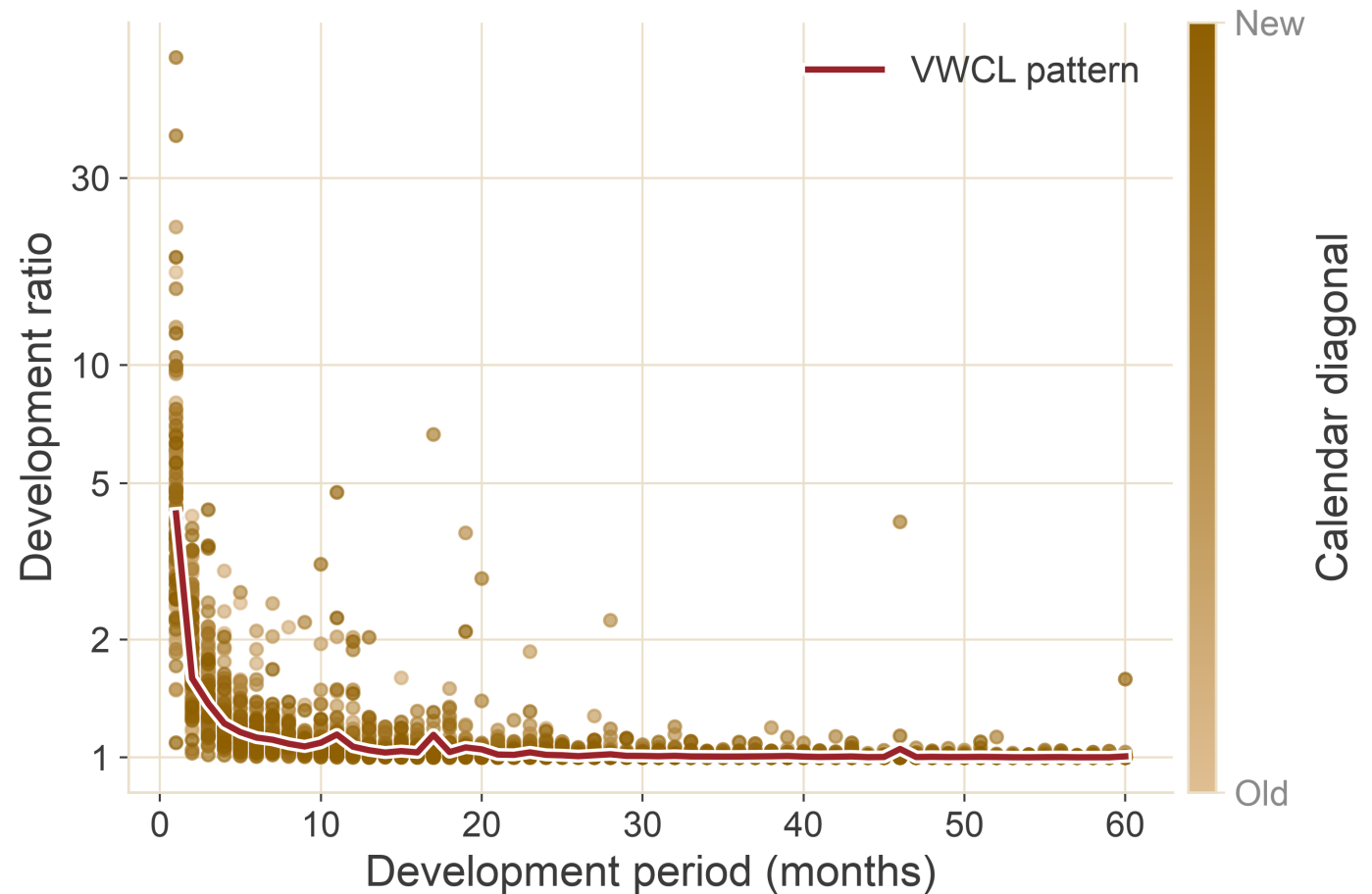
The chain ladder is already a fitted model

- Parameters: one per development period, f_j

$$\hat{C}_{i,j+1} = \hat{C}_{i,j} f_j$$

- Volume-weighted average minimises a loss function (Mack 1993)

$$L(f) = \sum_{i=1}^{I-1} \sum_{j=1}^n C_{i,j} (f_{i,j} - f_j)^2$$

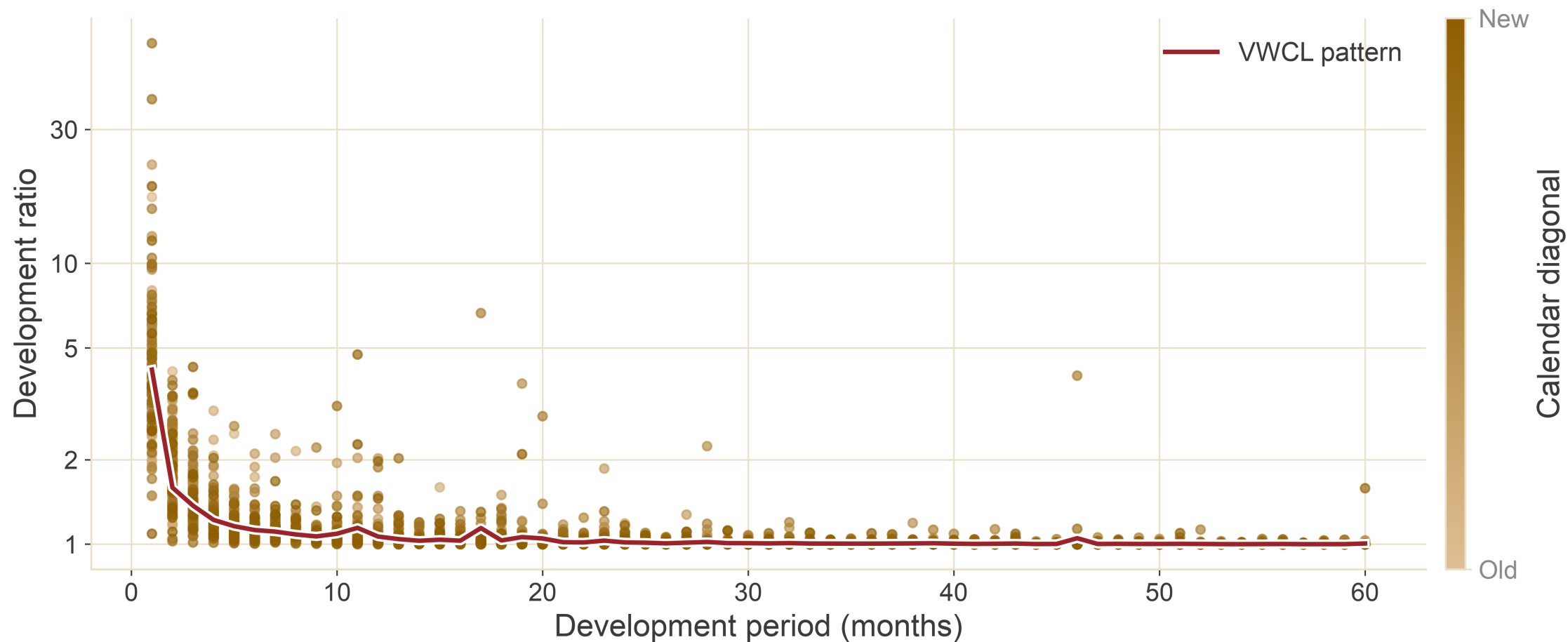


Two modelling mindsets

- Statistical modelling
- The world as probability distributions
- Aim to replicate the data generating mechanism
- Fit, then adjust by hand towards reality

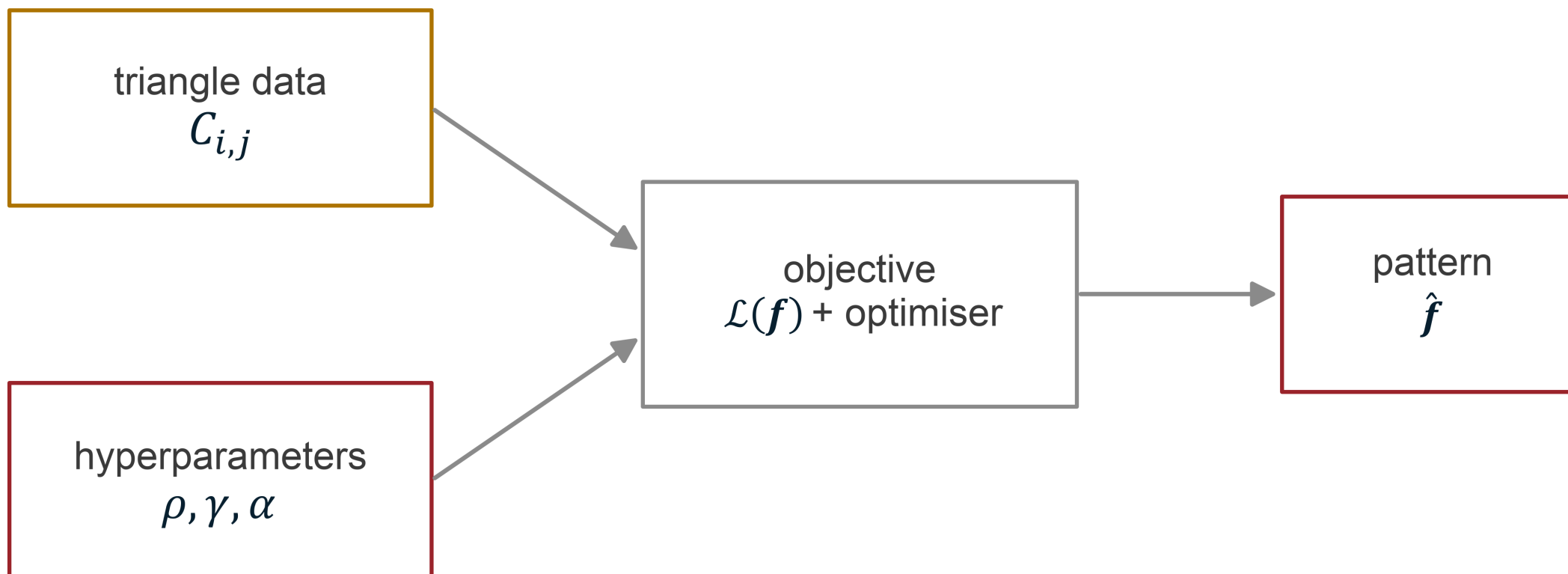
- Supervised learning
- Find the function with the lowest prediction error
- Held-out prediction decides, not the modeller
- Overfitting is mitigated by monitoring performance and frequent retraining

Rarely book an unadjusted fit in practice



Machine learning first principles on the chain ladder

Need more model flexibility, declared objectives beyond the current fit

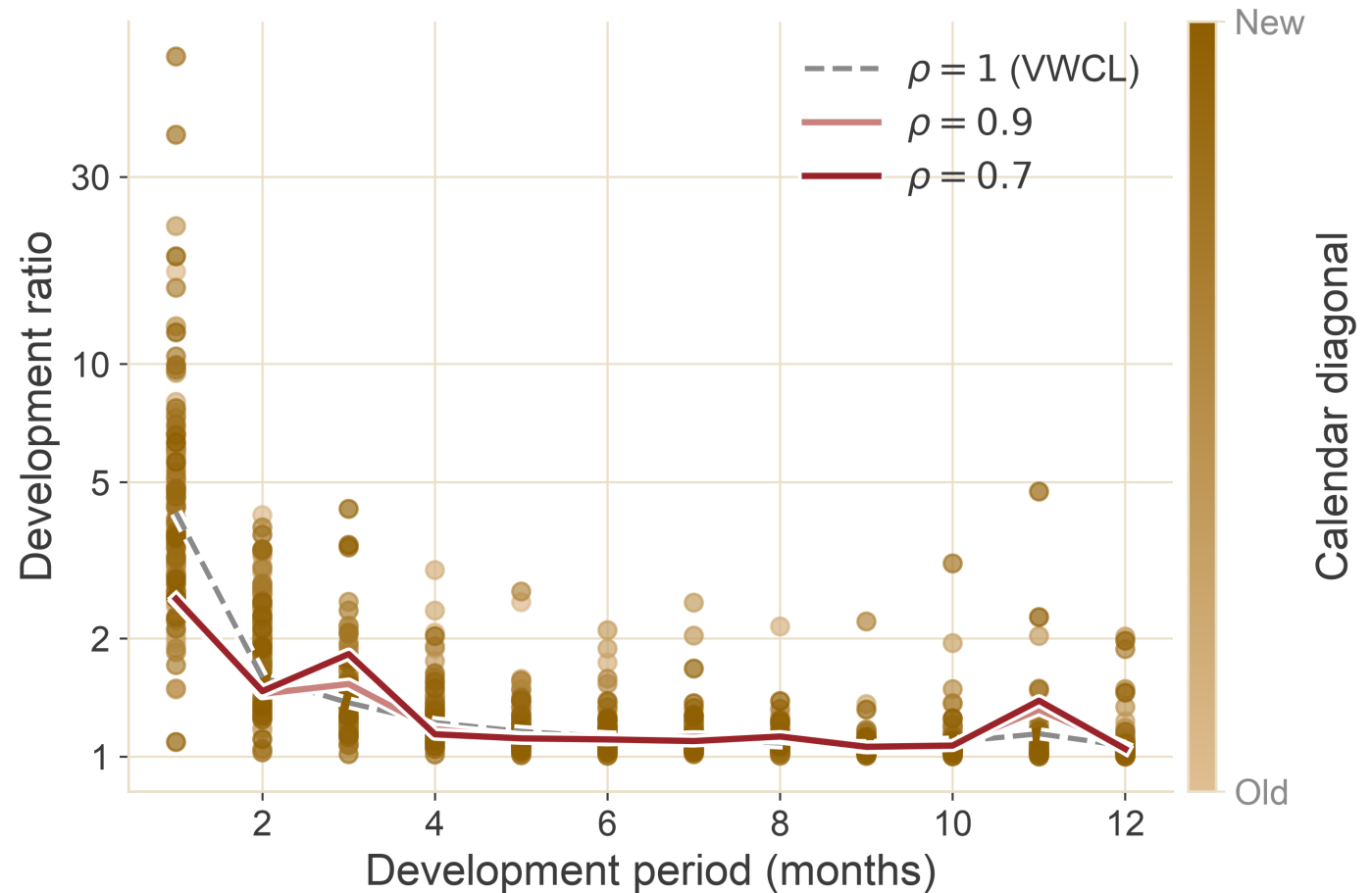


Trust recent experience more

- Weight recent calendar diagonals
- $\rho = 1$: normal Chain Ladder

$$w_{i,j} = \rho^{I-i-j} C_{i,j}$$

$$\text{Err}(\mathbf{f}; \rho) = \sum_{j=1}^n \sum_{i=1}^{I-j} w_{i,j} (f_{i,j} - f_j)^2$$

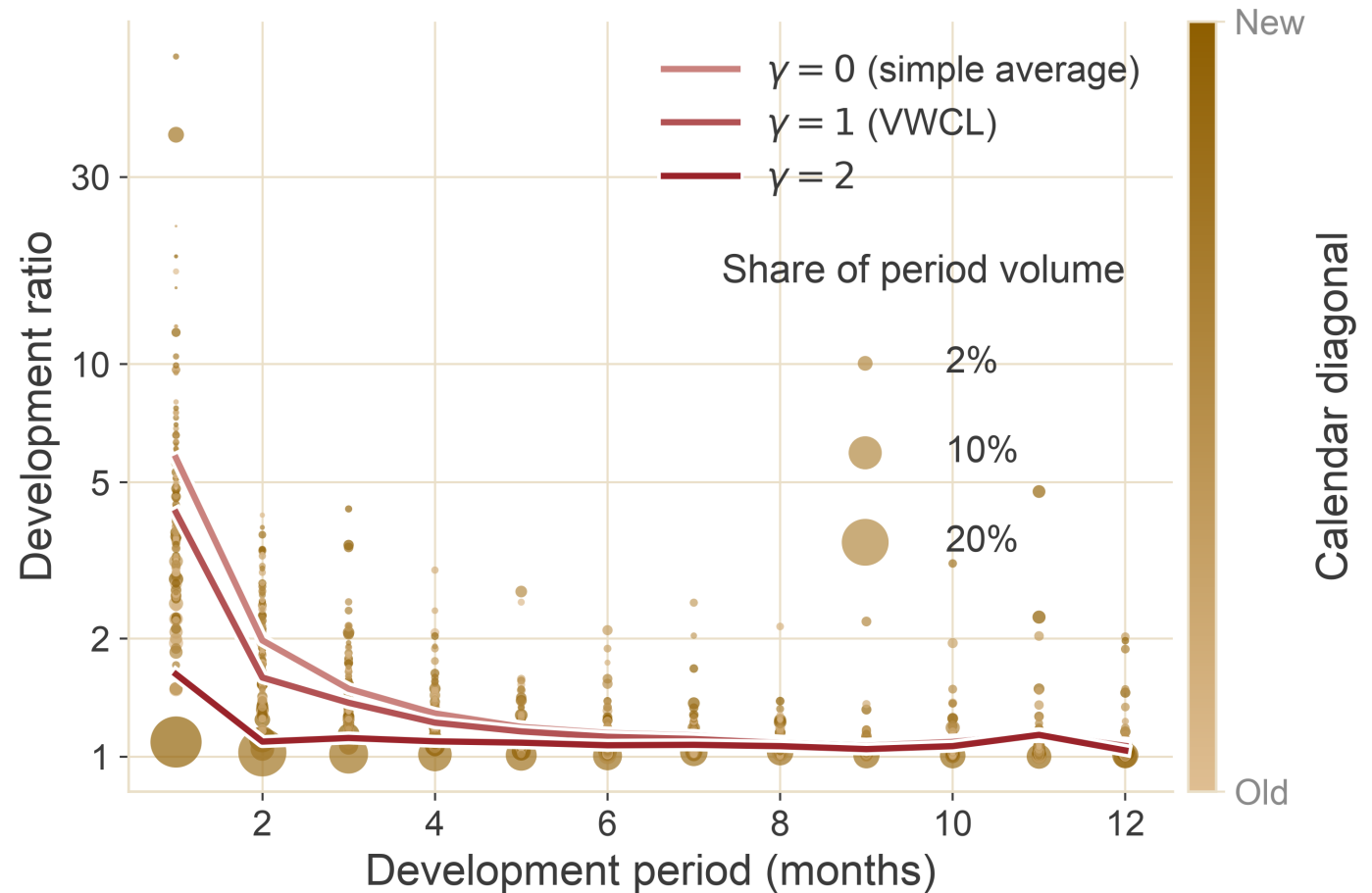


Control weights by claim size

- $\gamma = 0$ simple average
- $\gamma = 1$ volume-weighted
- $\gamma = 2$ unweighted squared error

$$w_{i,j} = \rho^{I-i-j} C_{i,j}^\gamma$$

$$\text{Err}(\mathbf{f}; \rho) = \sum_{j=1}^n \sum_{i=1}^{I-j} w_{i,j} (f_{i,j} - f_j)^2$$

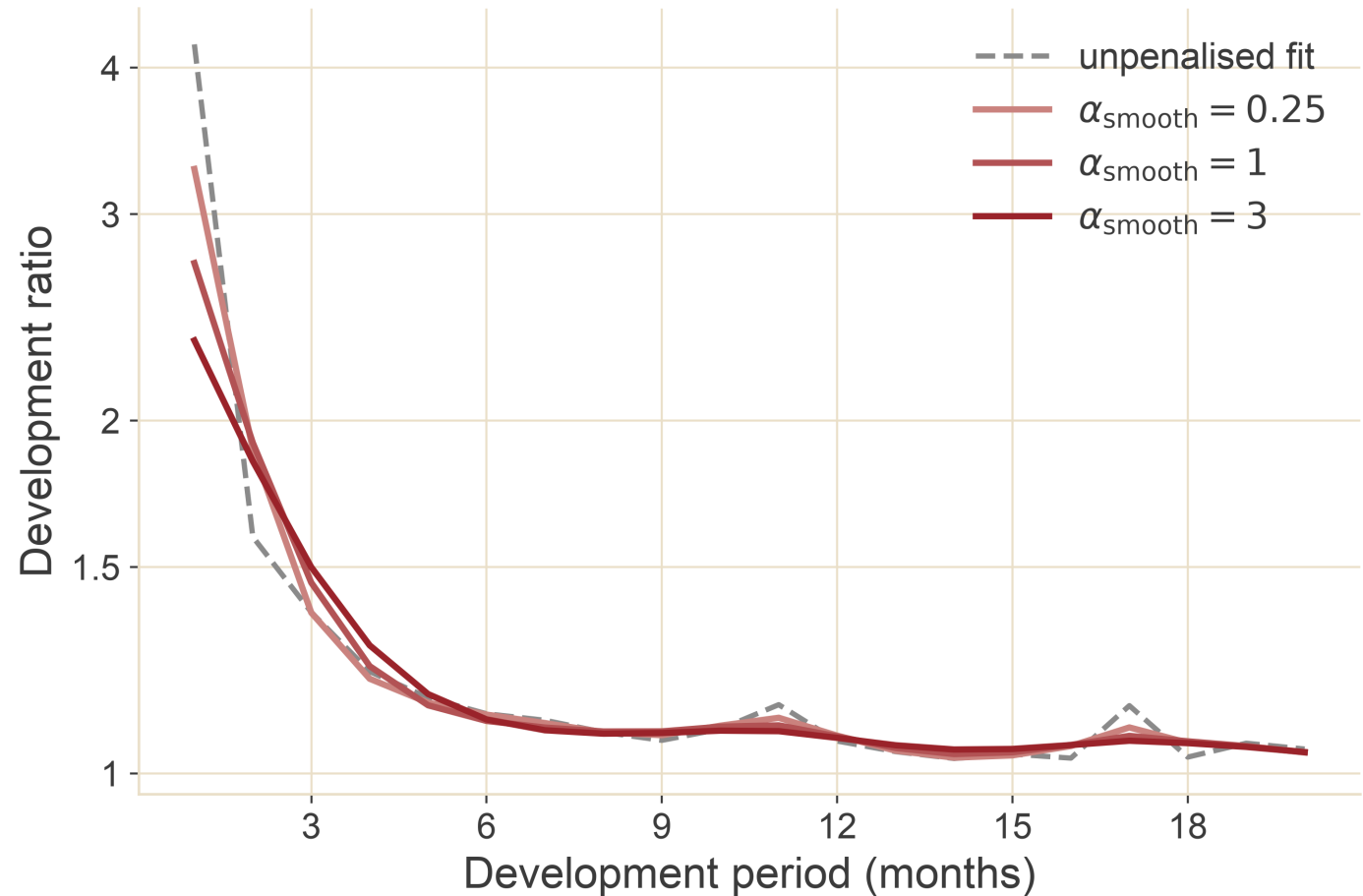


Smooth patterns balanced with goodness of fit

- Replaces: curve fitting, outlier adjustments

$$J_s(\mathbf{f}) = \sum_{j=2}^{n-1} w_j (2f_j - f_{j-1} - f_{j+1})^2$$

$$\mathcal{L}(\mathbf{f}) = \text{Err}(\mathbf{f}; \rho, \gamma) + \alpha_s J_s(\mathbf{f})$$

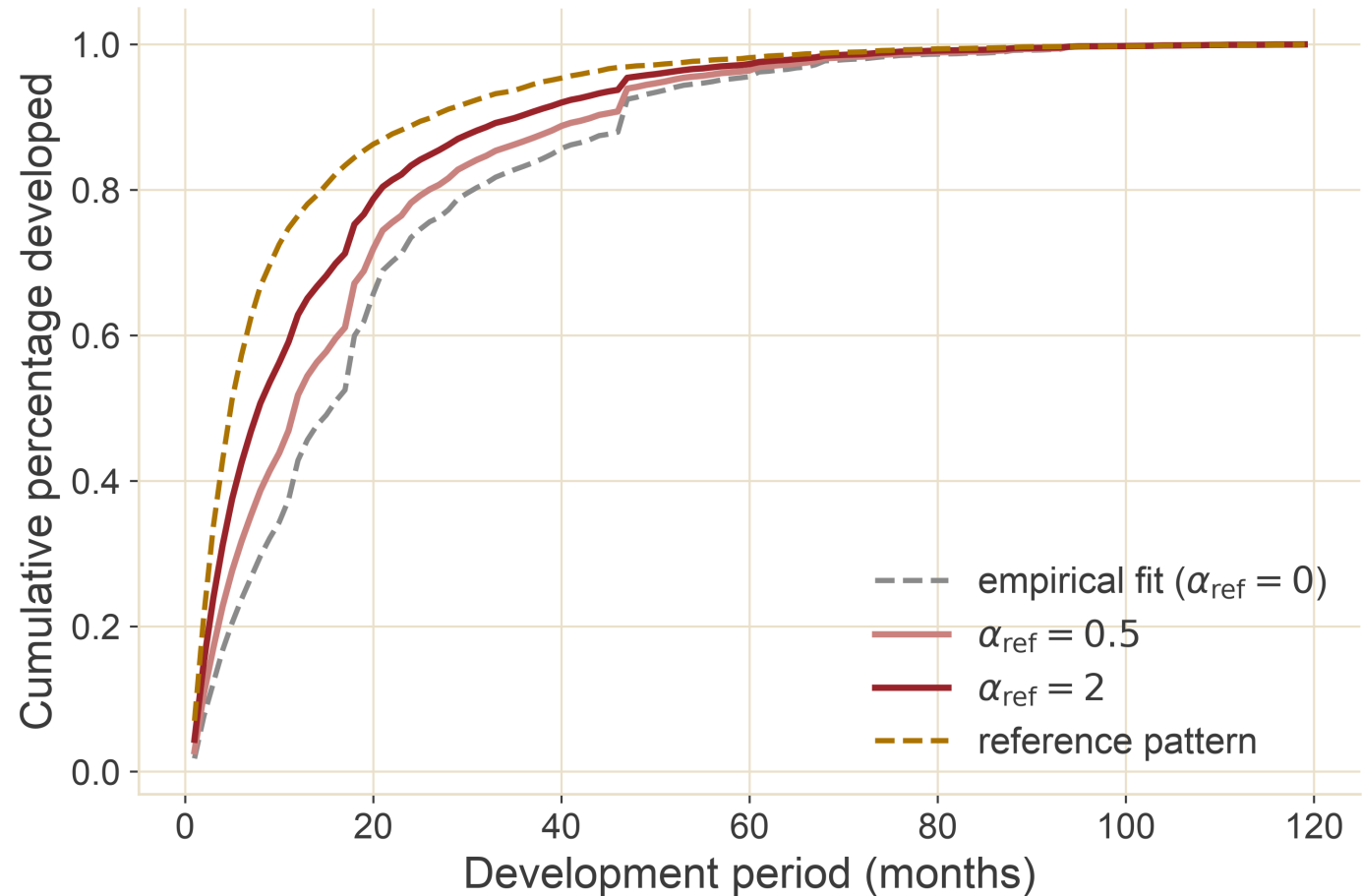


Pull towards a benchmark - and get credibility back

- Shaping towards a-priori reference reference
- Where J_{ref} takes a weighted squared difference between the patterns
- Interpretable credibility blend:

$$\hat{f}_j = \frac{\bar{f}_j + \alpha_{\text{ref}} f_j^{\text{ref}}}{1 + \alpha_{\text{ref}}}$$

$$\mathcal{L}(f) = \text{Err}(f; \rho, \gamma) + \alpha_s J_s(f) + \alpha_{\text{ref}} J_{\text{ref}}(f)$$



The full objective

- Each term is one declared objective as a penalty; weight says how much it matters
- Strictly convex - one closed form solution
- Judgement now lives in a handful of declarable weights
- Closed form:

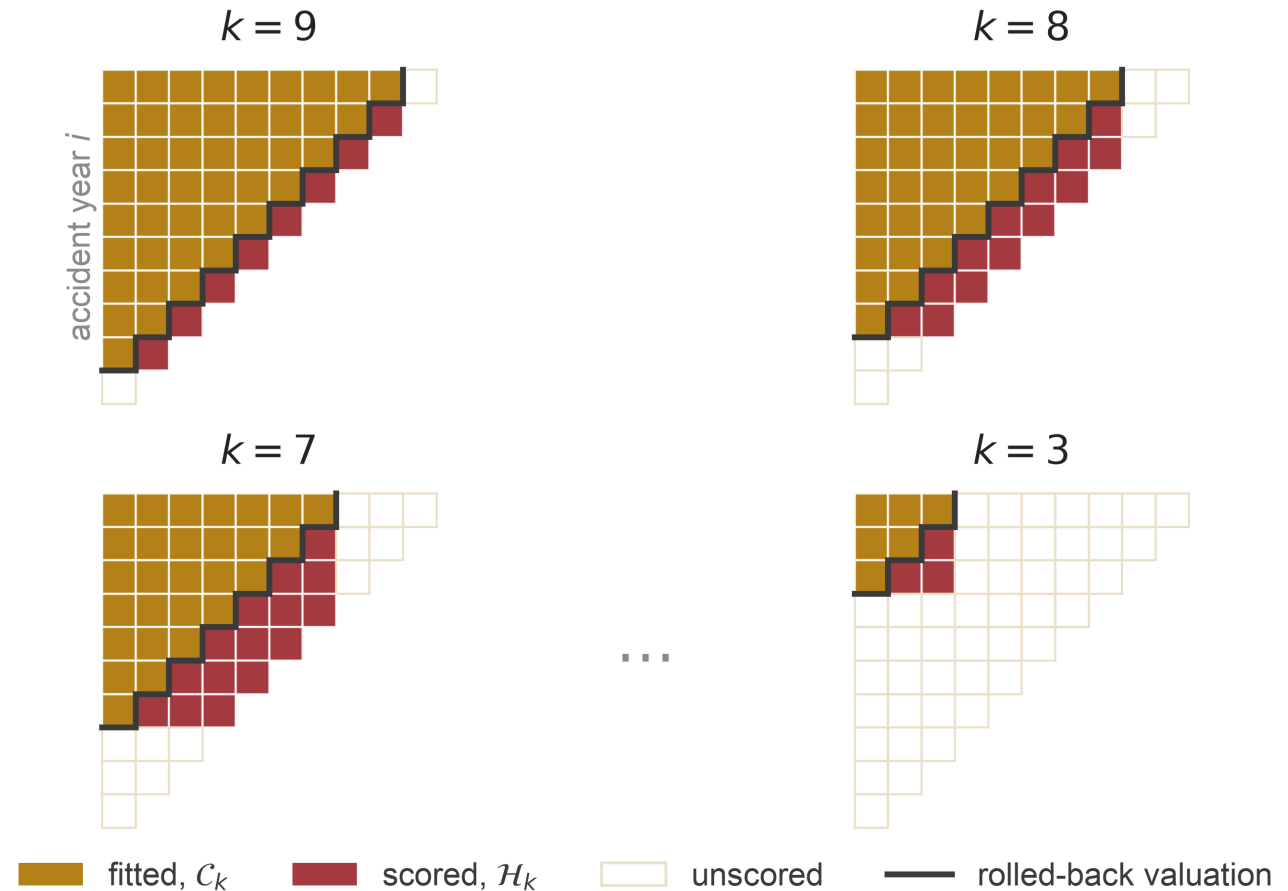
$$\left[(1 + \alpha_{\text{ref}})W + \alpha_s D^T \tilde{W} D \right] \hat{\mathbf{f}} = W(\bar{\mathbf{f}} + \alpha_{\text{ref}} \mathbf{f}^{\text{ref}})$$

$$\mathcal{L}(\mathbf{f}) = \text{Err}(\mathbf{f}; \rho, \gamma) + \alpha_s J_s(\mathbf{f}) + \alpha_{\text{ref}} J_{\text{ref}}(\mathbf{f})$$

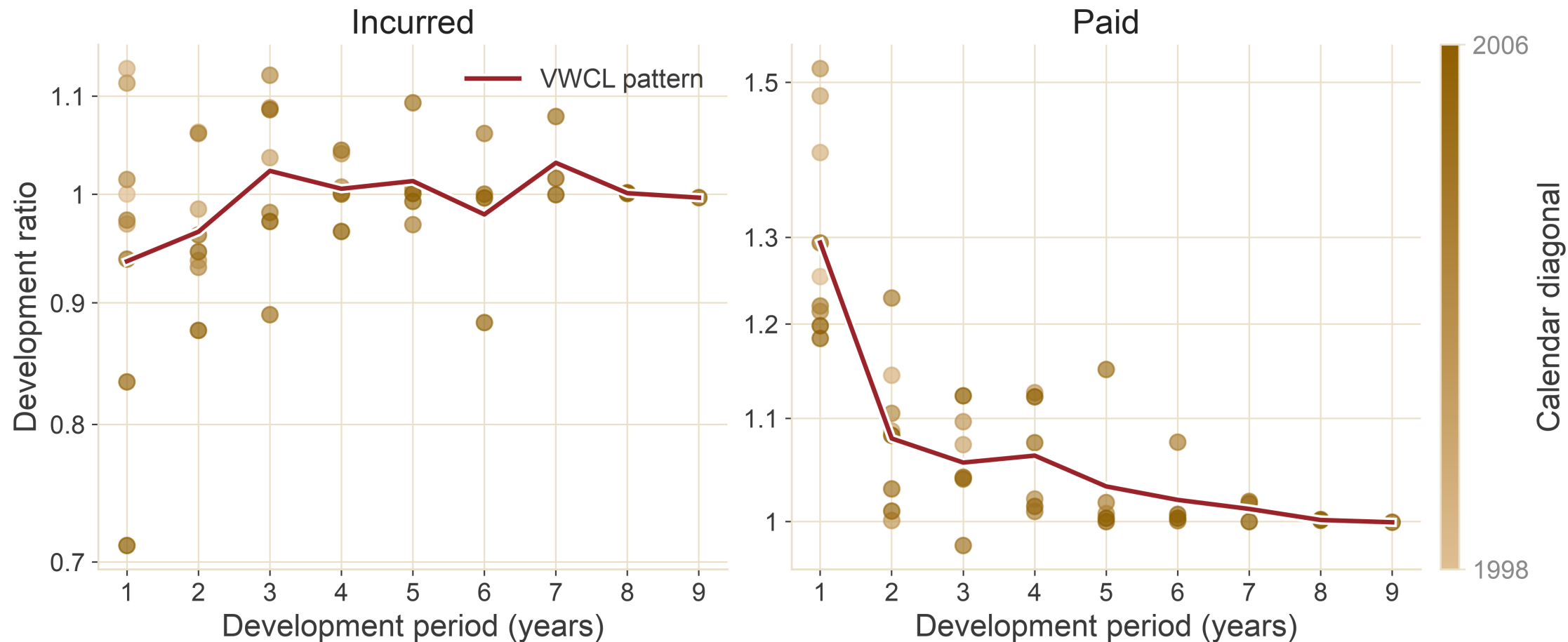
Train the weights on held-out diagonals

- Score one-step errors
- Standardised by Mack's variance
- Grid search hyperparameters

$$\text{score}_k = \frac{1}{|\mathcal{H}_k|} \sum_{(i,j) \in \mathcal{H}_k} \frac{(C_{i,j+1} - C_{i,j} \hat{f}_j^{(k)})^2}{\sigma_j^2 C_{i,j}}$$

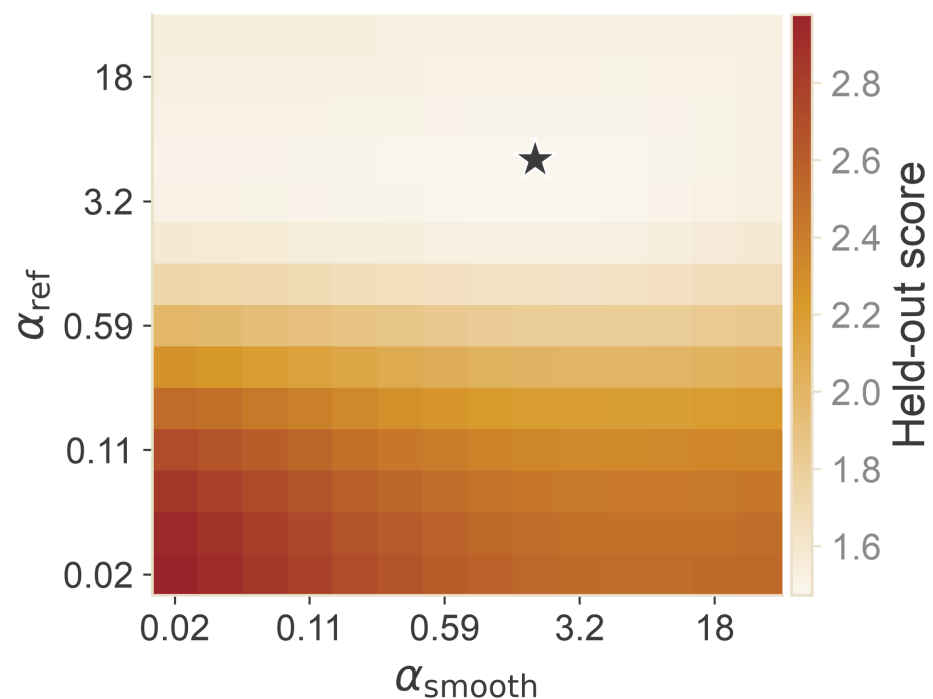


Example: small insurer, noisy other liability class

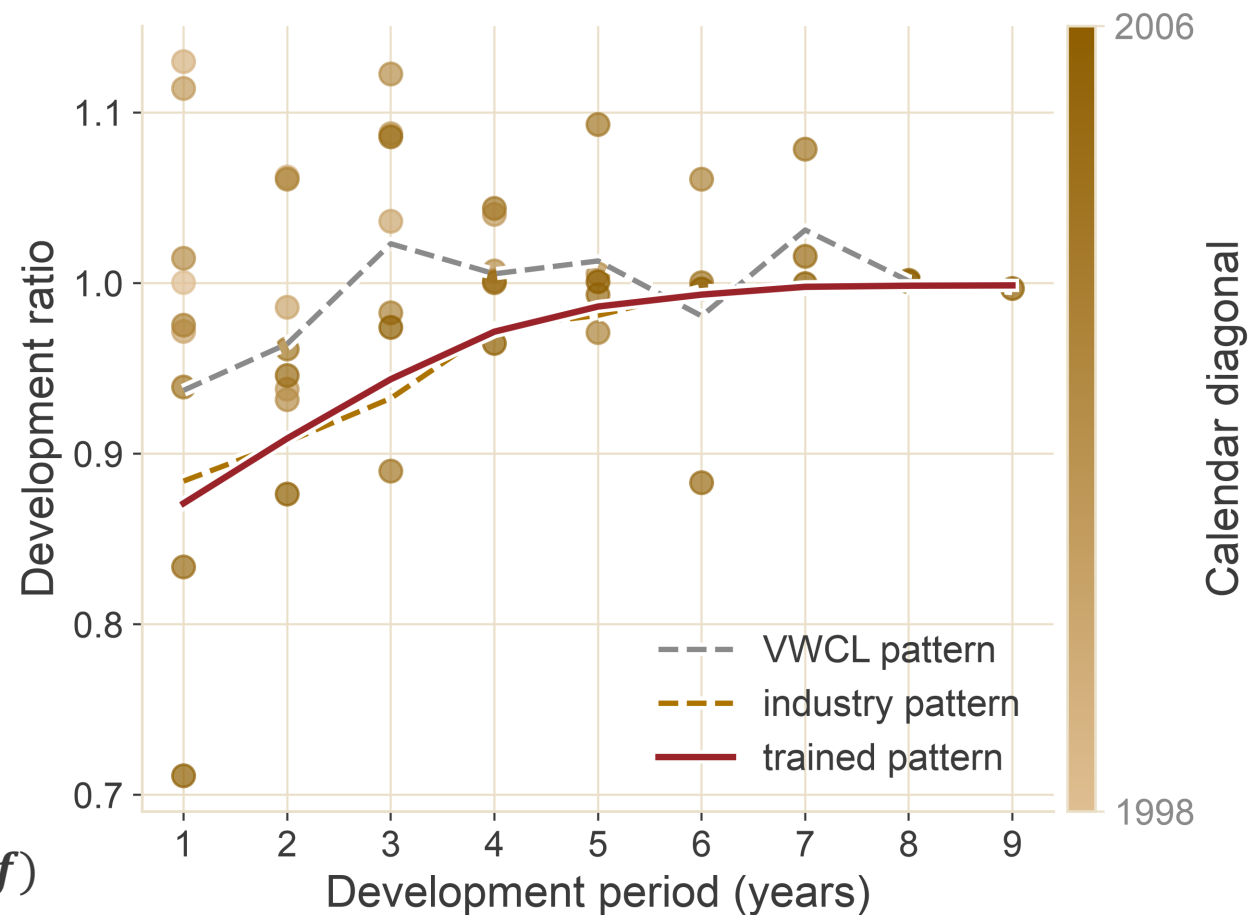


Example: Train the incurred chain ladder

Industry pattern as reference pattern



$$\mathcal{L}(f) = \text{Err}(f; 0.50, 1) + 1.84 J_s(f) + 5.68 J_{\text{ref}}(f)$$

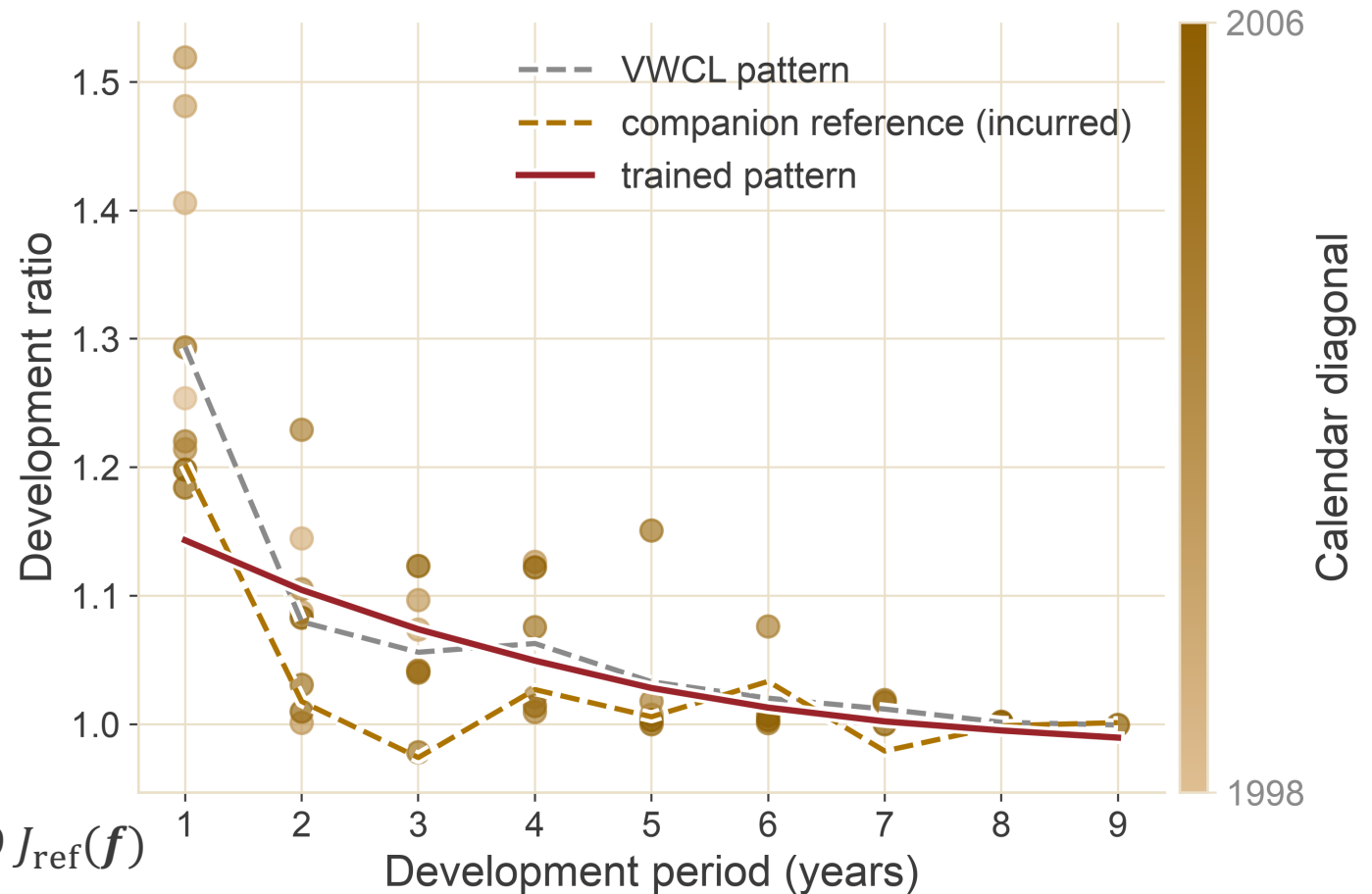


Example: Training the paid chain ladder

- Companion reference pattern derived derived from incurred

$$f_j^{\text{ref}} = \bar{f}_j^P \cdot (\hat{f}_j^I / \bar{f}_j^I)$$

$$\mathcal{L}(f) = \text{Err}(f; 0.15, 1) + 10.0 J_s(f) + 0.59 J_{\text{ref}}(f)$$



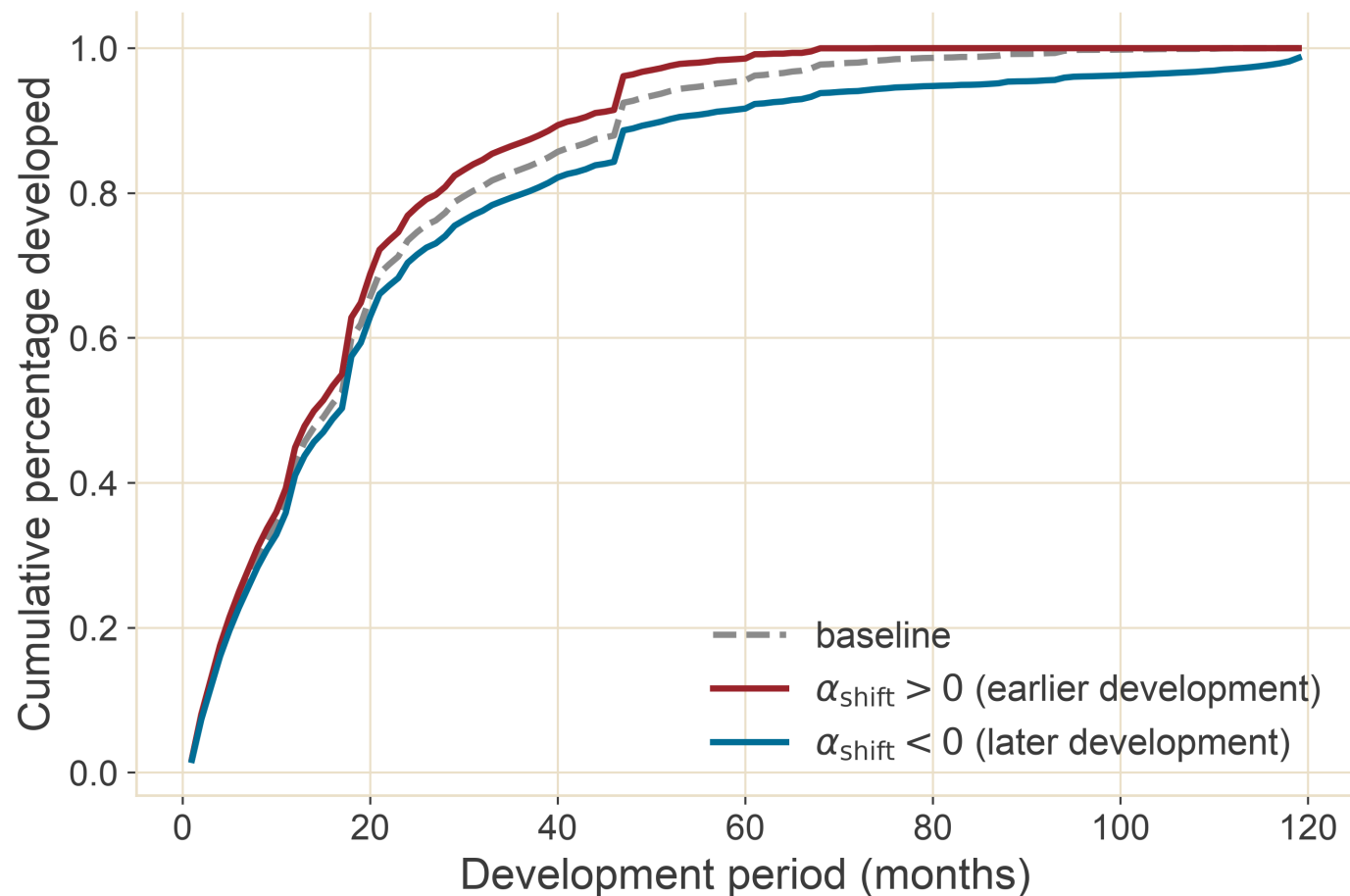
Adjusting for information not in the data

- Experience adjustment
- Changes how observed experience is read
- Testable on history
- Can be trained
- Can also be declared to reduce overfitting

- Prospective adjustment
- The future differs from the past
- No validation data can confirm it
- This adjustment scores poorly on the past
- Must be declared

Control weight in the tail

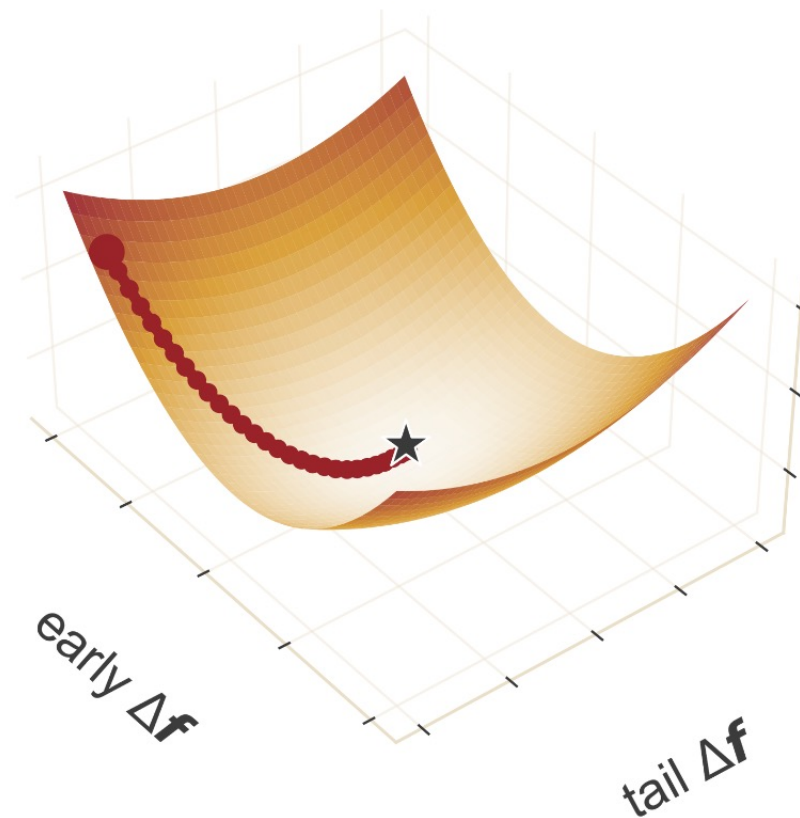
- Not convex
- No closed form solution



$$\mathcal{L}(f) = \text{Err}(f; \rho, \gamma) + \alpha_{\text{shift}} \sigma(\hat{\beta}^*(f))$$

Solve a non-convex objective with no closed-form

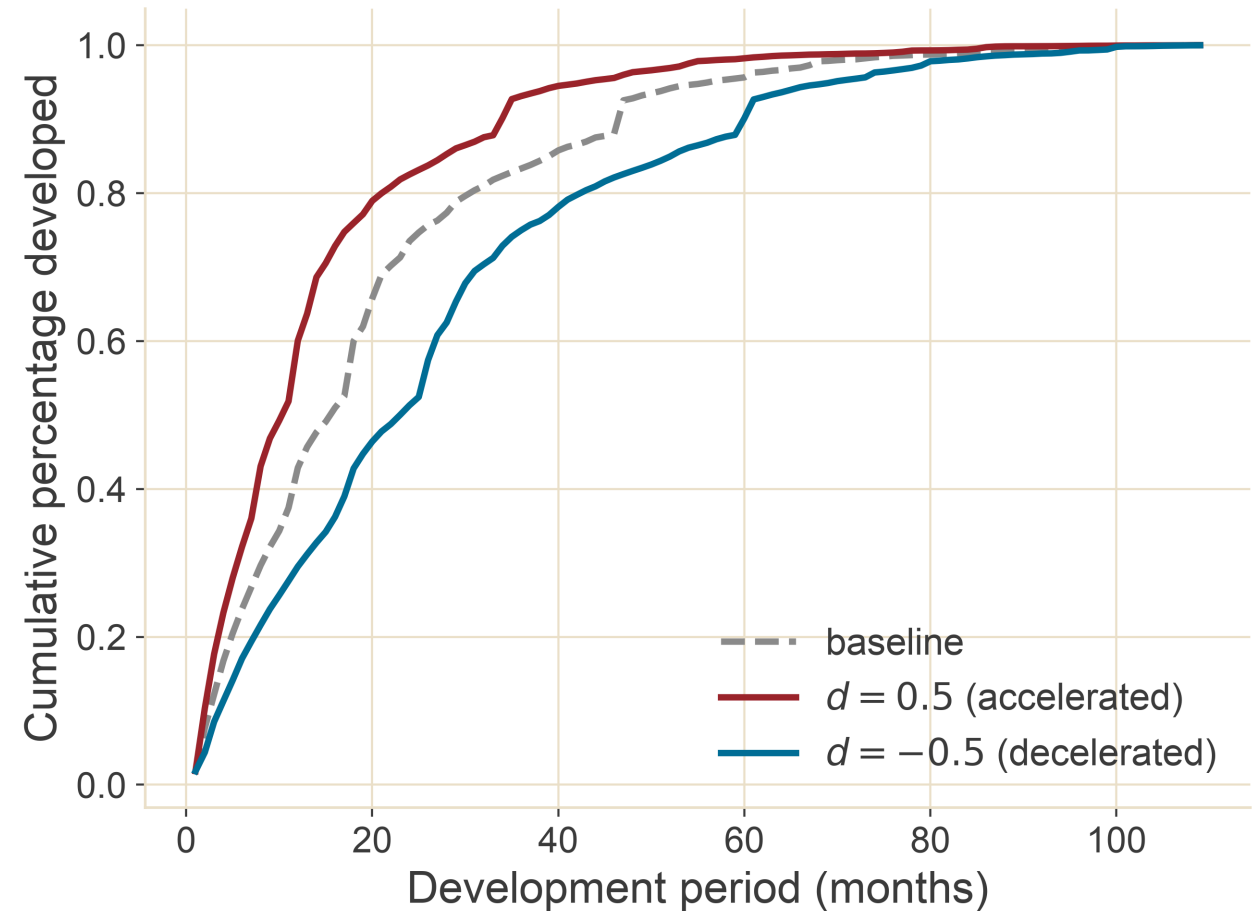
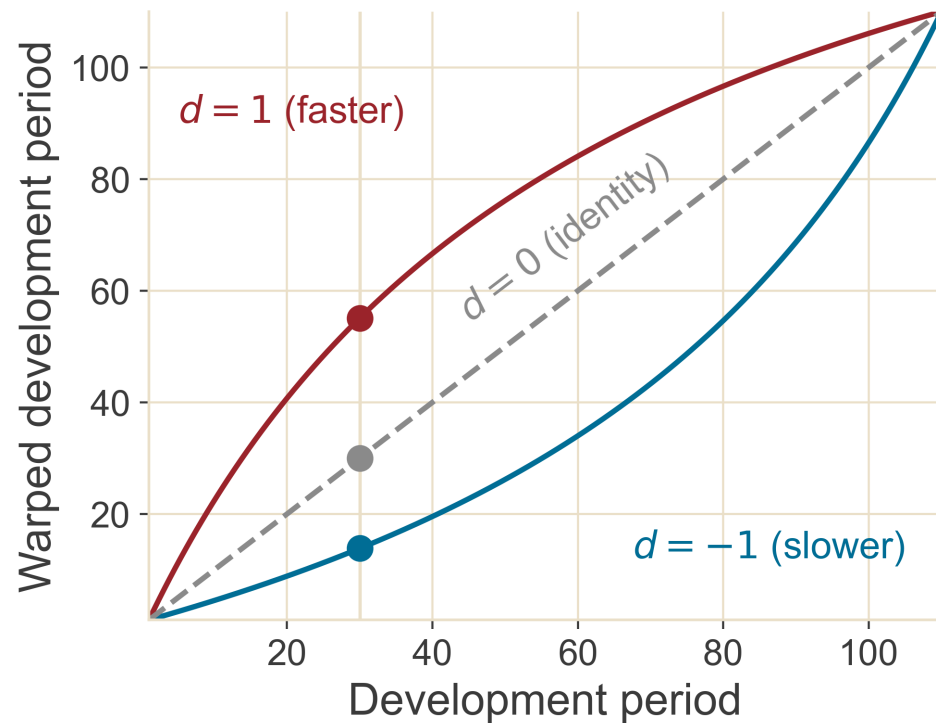
- A numerical method is required
- Differentiable almost everywhere
- Gradient-based optimiser makes sense



objective $\mathcal{L}$

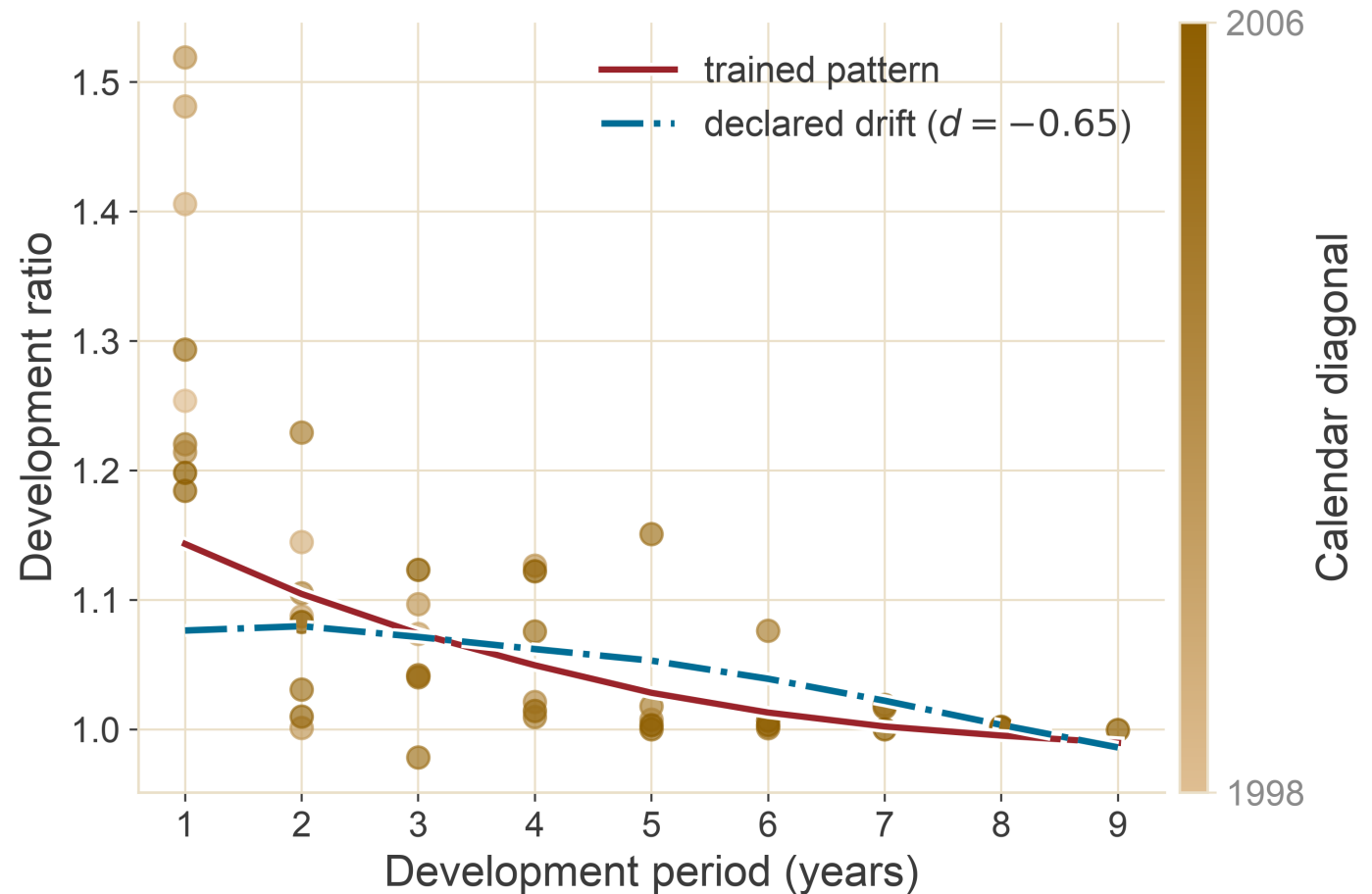
Pattern drift: re-time the whole pattern

Let $f_j^* = \psi_d(\hat{f}_j)$, where d warps j



Example: A declared prospective adjustment

- Hypothetical settlement delay
- The triangle cannot yet show it
- Hyperparameter weight declared, not trained
- Drift strength $d = -0.65$



Example: Did it predict actual experience well?

- Reserve errors relative to the actual run-off of 2,915
- Trained paid accuracy is closest (spurious)
- For illustration, not empirical performance claim

Projection	Total ultimate (\$000)	Reserve (\$000)	Reserve error
VWCL incurred	41,541	5,458	+87.2%
Trained incurred	38,083	2,000	-31.4%
VWCL paid	40,374	4,291	+47.2%
Trained paid	38,967	2,884	-1.1%
Declared prospective	40,621	4,538	+55.7%
Actual run-off	38,998	2,915	-

Supervising the Chain Ladder

Manual adjustment



Objective term

Trust the data	exclude ratios, weight recency	→	$\text{Err}(\mathbf{f}; \rho, \gamma)$
Smooth	fit a curve by eye	→	$\alpha_{\text{smooth}} J_{\text{smooth}}$
Benchmark	pull towards industry	→	$\alpha_{\text{ref}} J_{\text{ref}}$
Re-shape	edit many ratios at once	→	$\alpha_{\text{shift}} J_{\text{shift}}$

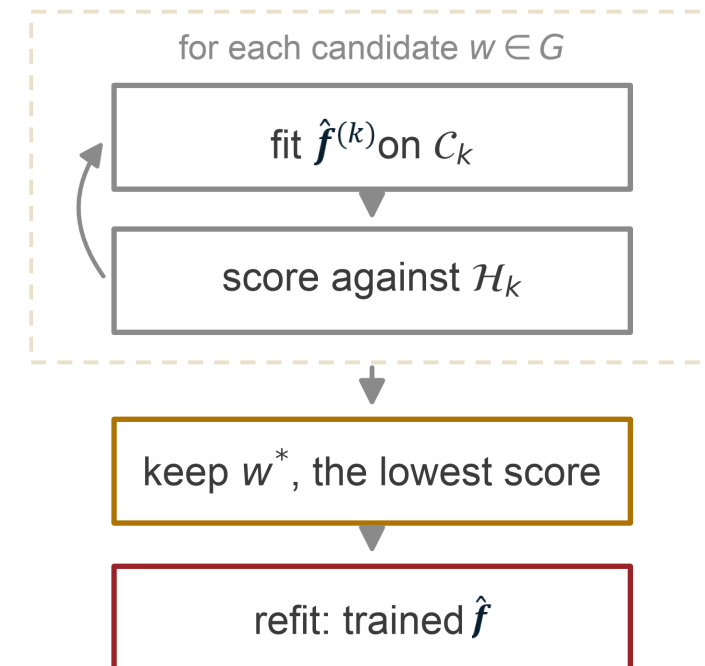
Experience	what the history shows	→	trained
Prospective	what the history cannot show	→	declared

$$\mathcal{L}(\mathbf{f}) = \text{Err}(\mathbf{f}; \rho, \gamma) + \sum_{p \in \text{Pen}} \alpha_p J_p(\mathbf{f})$$

the fitted pattern is its minimiser

Training loop

hold out each recent diagonal



Questions

james.grove@dyna-mo.com
anesu.shuro@dyna-mo.com

#TASKConvention2026

When the framework fails (for anticipated questions)

- Small triangles: too few diagonals to support the hold-out splits
- Genuinely unstable books: no weighting rescues an unrepresentative history
- Errors or misleading information still lead to overfitting and poor generalisation
- Declared objectives may conflict - opposing pulls dilute the empirical fit
- The chain ladder's own limitations still apply