The Problem Statement.

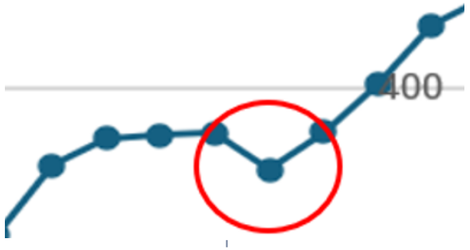
How does a model-based valuation change with changing economic and actuarial parameters?

And what do we do knowing that all models have errors somewhere, while covering a double-digit number of risks?

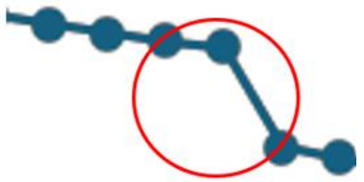
Actuary

Lets extract model patterns and check problematic ones

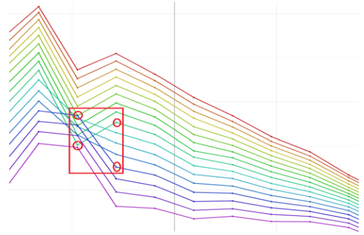
PVFP = F(Equity Index)



PVFP = F(Longevity)



PVFP = F(Rates, Longevity)



Triage findings: Bug, Artefact or Learning?

If bug or artefact: Which model to improve and how to do so?

Carry out model upgrade

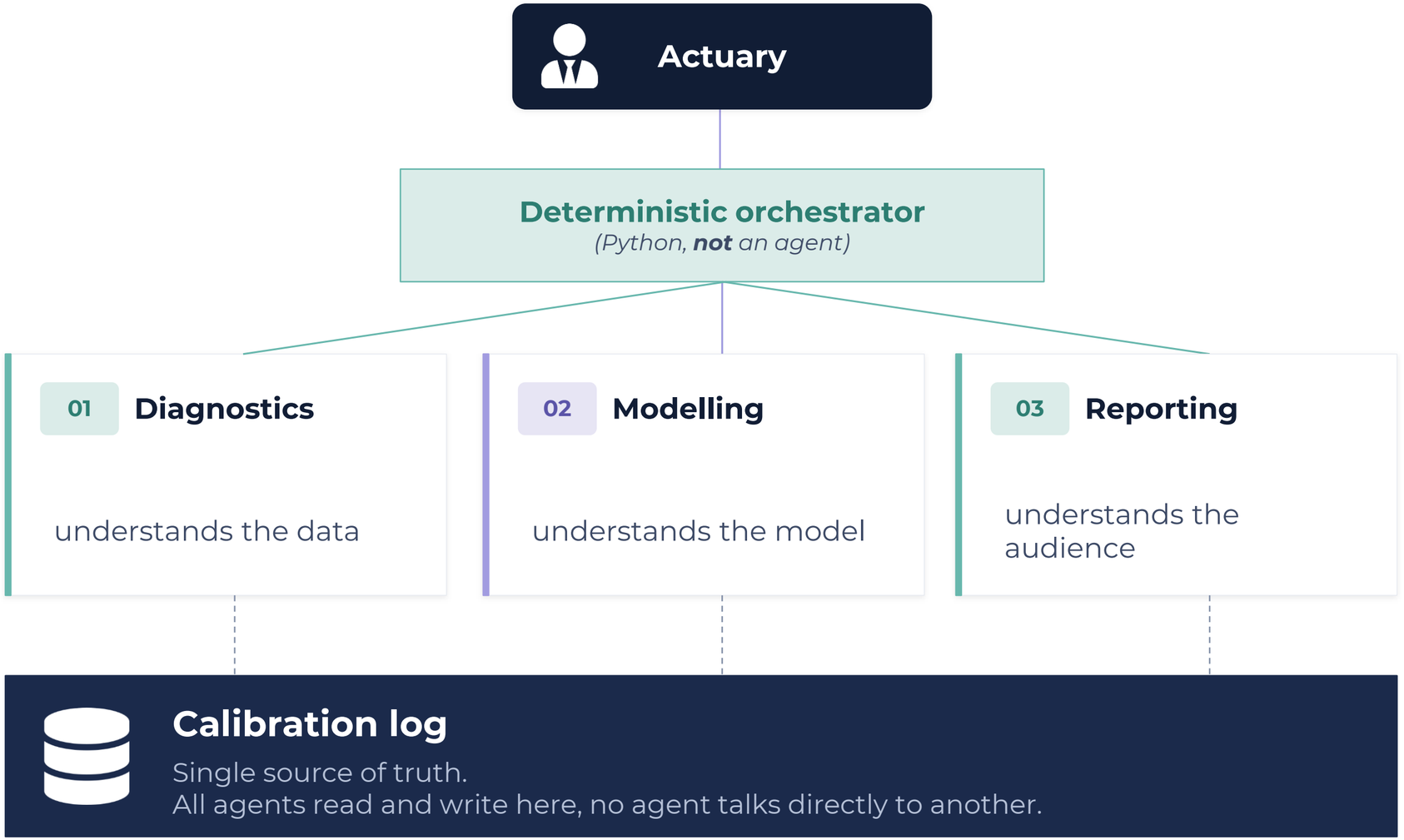
Document change, impact and sign-offs.

The Agentic Team.

Three specialist agents, coordinated by a deterministic orchestrator. They communicate through one shared file, not through chat.

3 + 1

specialists + orchestrator



The team is small on purpose.

Six judgement domains would be a committee.

KYNESIS

Three Decision Gates.

HUMAN IN THE LOOP

The actuary stays in charge but is interrupted only at three checkpoints, not at every step. Everything else runs autonomously.

Where the actuary makes the decision

A

Diagnostics confirmation

"Is each flagged kink real or an artefact?"

Routes the workflow: fix at source, perform ML enhancement, or proceed clean.

B

Fit acceptance

"Are residuals acceptable?"

Approve the fit or request a re-run with adjusted parameters.

C

Report sign-off

"Am I willing to put my name on this?"

Final sign-off recorded with hash and timestamp; report locks for the regulator.

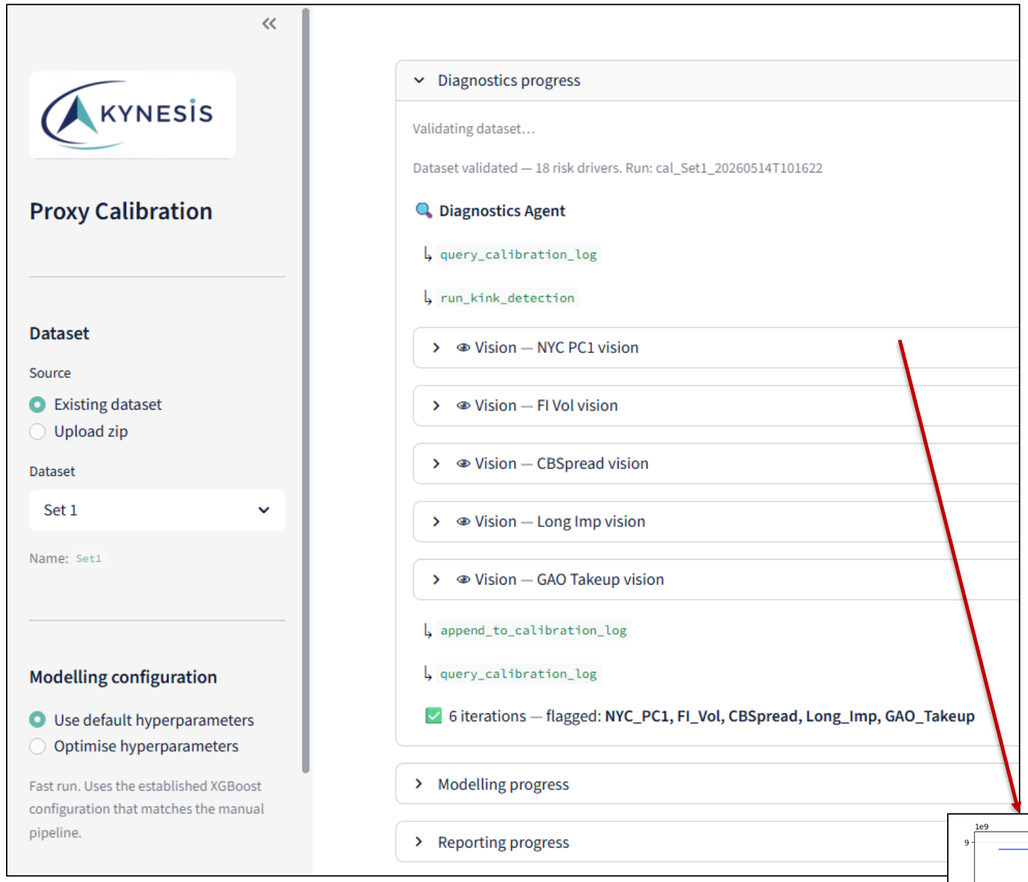
Actuary removed from setting-up proxy model calibration, review, manual debug of loss functions

To apply quality judgement at key decision gates.

AI Agent
Detecting Issues.

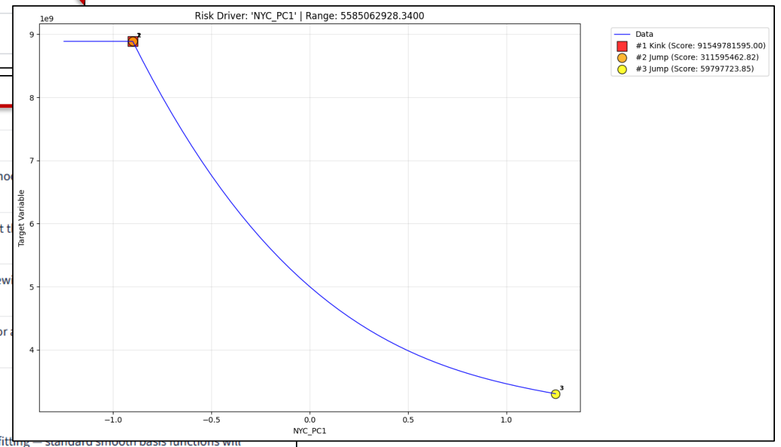
AI Agent cutting through the cashflow model looking for odd behaviours.

Background #1: Detection of Cashflow Model artefacts



Suspicious behaviours of the underlying cashflow projection model being found, such as discontinuities, kinks, cliff edges, inflextion points etc.

Kink Detected!



Material Issues for Investigation

Issue	Detail
Gradient discontinuity at x ≈ −0.90	The sharp kink is algorithmically and visually confirmed. A proxy function with a continuous basis (e.g., polynomial, neural network with smo feature without explicit accommodation.
Plateau saturation (x < −0.90)	The flat region implies the proxy must correctly capture a censored or bounded response. If the fitting region does not adequately represent t stress scenarios where NYC_PC1 < −0.90, which could be material for tail capital metrics (e.g., 1-in-200 SCR).
Basis function adequacy	Recommend verifying whether the proxy architecture includes a **kink basis term** (e.g., a hinge/ReLU-type term centred at x ≈ −0.90) or piecew neighbourhood of the kink are likely to be systematically biased.
Economic interpretation	The kink should be traced to a specific modelling feature in the full cashflow model — likely a non-linear option payoff, a contractual floor, or a specific market level. This should be documented in the calibration report.

Overall Assessment

The kink flag is valid and material. The transition at x ≈ −0.90 represents a genuine non-smoothness in the response surface. This driver requires special treatment in proxy fitting — standard smooth basis functions will produce a poor local fit at this transition point, with likely downstream impact on goodness-of-fit metrics and capital estimates under stressed scenarios.

AI Agent
Detecting Issues.

AI Agent cutting through the oroxy model looking for odd behaviours.

Background #2: Detection of Proxy Model Artefacts

1. Flagged / Kink Drivers — Detailed Assessment

NYC_PC1 *(High severity — confirmed misfit)*

The actual cashflow profile exhibits a sharp convex kink in the region x ≈ −0.3 to 0.0, transitioning from a steep decline to a near-flat plateau. The degree-4 polynomial tracks this reasonably in the central region but **over-smooths the inflection point**, producing visible residuals on the left-hand shoulder (x ≈ −0.8 to −0.3) where the fitted curve sits above the actuals. The right-hand tail (x > 0.5) shows modest positive bias. This is a classic polynomial underfitting pattern at a kink location — the polynomial is forced to allocate curvature globally, compromising local fit at the transition. **A piecewise polynomial or spline basis is indicated.**

Long_Imp *(High severity — confirmed misfit)*

This driver shows the most visually pronounced misfit of all 18 factors. The actual profile is strongly asymmetric: nearly flat for x < −0.3, then dropping sharply toward the right tail. The polynomial fit **fails to capture the flat left-hand region**, producing a smooth S-shape that materially overestimates cashflows for x ∈ [−1.0, −0.3] and underestimates in the steep descent region x ∈ [0.0, 0.8]. The divergence in absolute terms appears to span approximately 100–150 units on the plotted scale, suggesting this driver alone may be a **primary contributor to the RMSE breach**. A knot at approximately x = −0.2 would be the natural intervention point.

GAO_Takeup *(High severity — confirmed misfit)*

The actual profile is highly non-linear: relatively stable across x ∈ [−1.0, 0.3], then declining sharply and flattening again at the right extreme (x > 0.8), producing an elongated reverse-J shape. The degree-4 polynomial **tracks the left-hand flat region adequately but significantly undershoots the right-hand descent**, with the fitted curve continuing to fall through the actuals at x > 0.8 where the actuals re-flatten. This terminal re-flattening is structurally un-fittable with a global polynomial without a very high degree. The residuals at the extreme right tail are visually among the largest in the entire chart set. **This driver warrants spline or segmented treatment.**

FI_Vol *(Flagged — moderate misfit)*

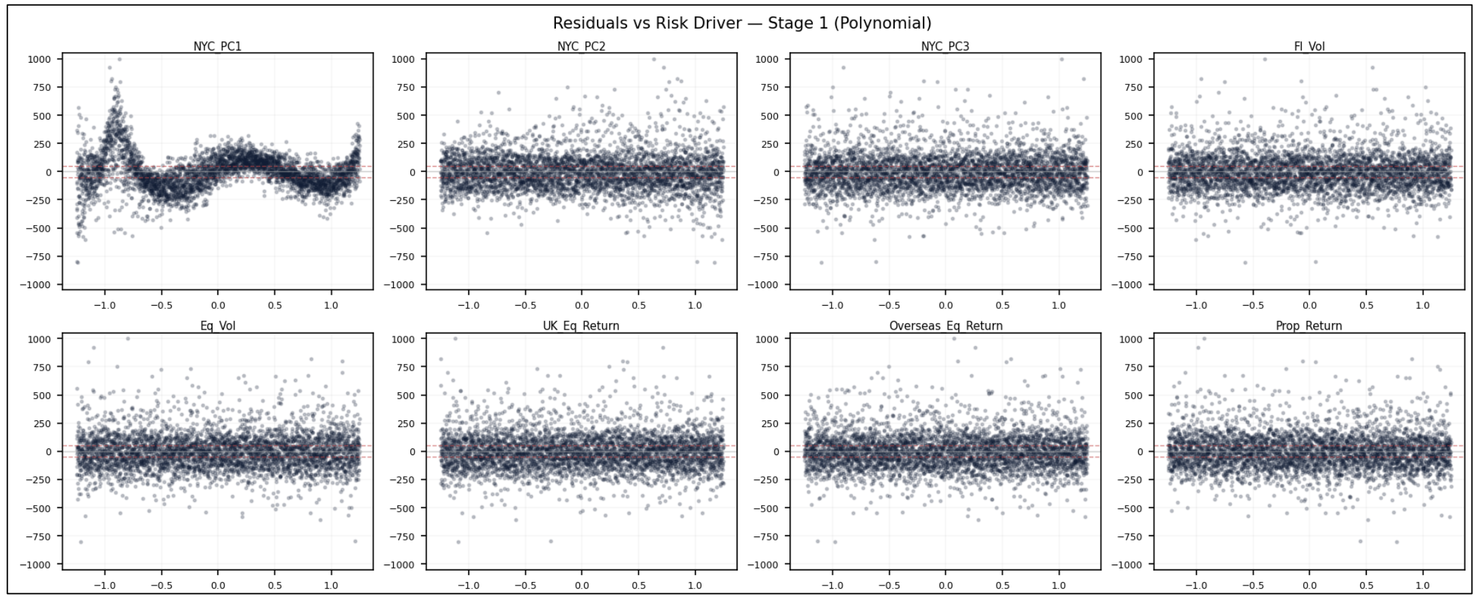
The actual profile is U-shaped with a shallow left arm and steeper right arm — a mildly asymmetric convex curve. The polynomial fit captures the general shape but **systematically underestimates across the left-to-central region** (x ∈ [−1.0, 0.5]), with the fitted curve sitting below actuals by a near-constant offset in this range before converging at the right tail. This suggests the polynomial is allocating too much curvature to the right-hand upturn at the cost of the left-hand plateau. A degree-4 term should theoretically accommodate this, but the global fitting constraint is introducing bias. The misfit is moderate in absolute scale but directionally consistent.

CBSpread *(Flagged — moderate to material misfit)*

The actual exhibits a sigmoid / logistic shape — increasing slowly from left, accelerating through mid-range, then flattening at the right. The polynomial fit produces an S-curve that **leads the actual profile on the left (overfit) and lags on the right (underfit)**, with the crossover near x = 0.0. The right-hand tail divergence is visually material. This is a well-known failure mode for polynomials fitting sigmoid-shaped responses — the basis cannot simultaneously fit the flat tails and the central slope without oscillation. **This driver is a candidate for logistic or spline transformation.**

Goodness of fit of an LSMC Proxy Model being analysed. Human in the loop: Actuary to decide whether or not a Machine Learning improvement is needed.

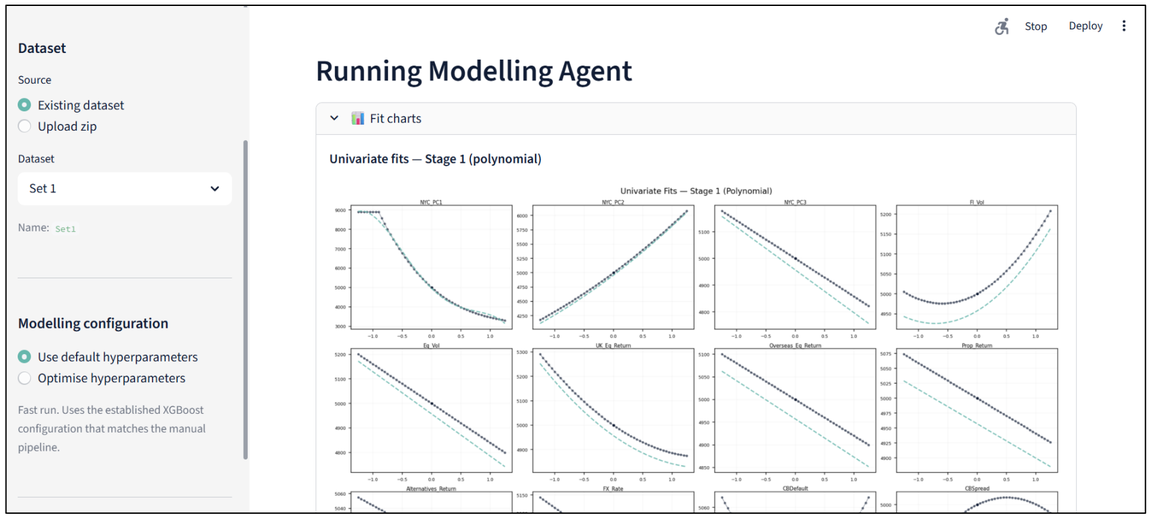
Initial fit shows residuals are heteroscedastic!



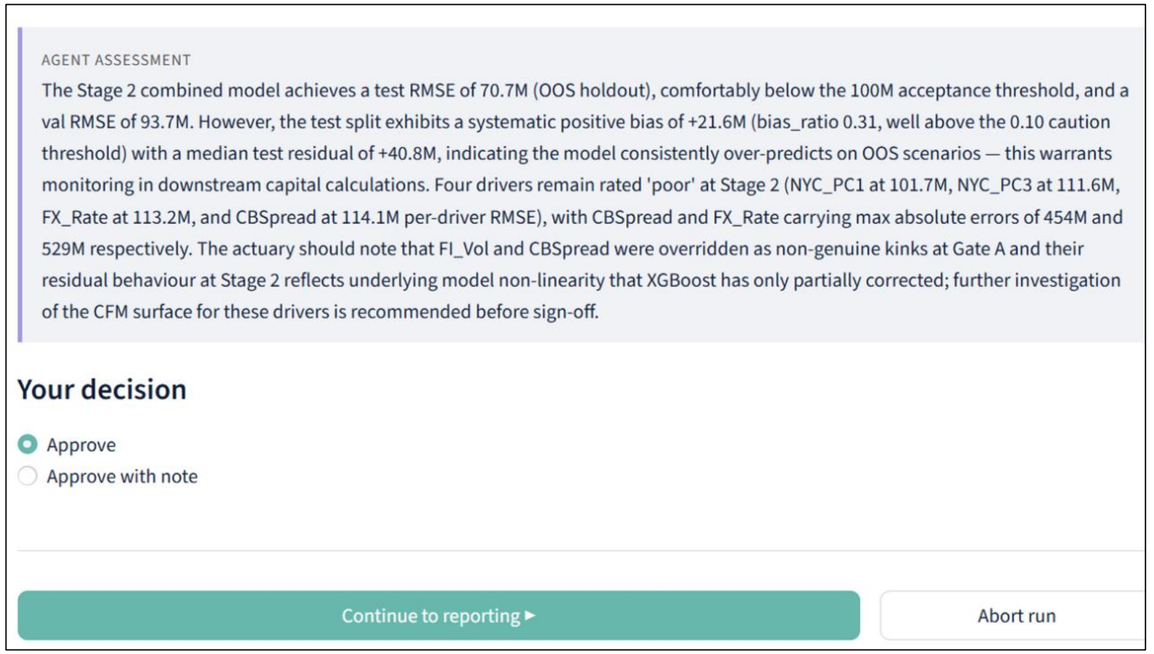
AI Agent
Upgrading Model.

AI Agent applying Machine Learning methods to improve a Proxy Model.

Background #3: Proxy Model Upgrade



Machine Learning methods being deployed to upgrade a traditional proxy. Improvements achieved being tracked.

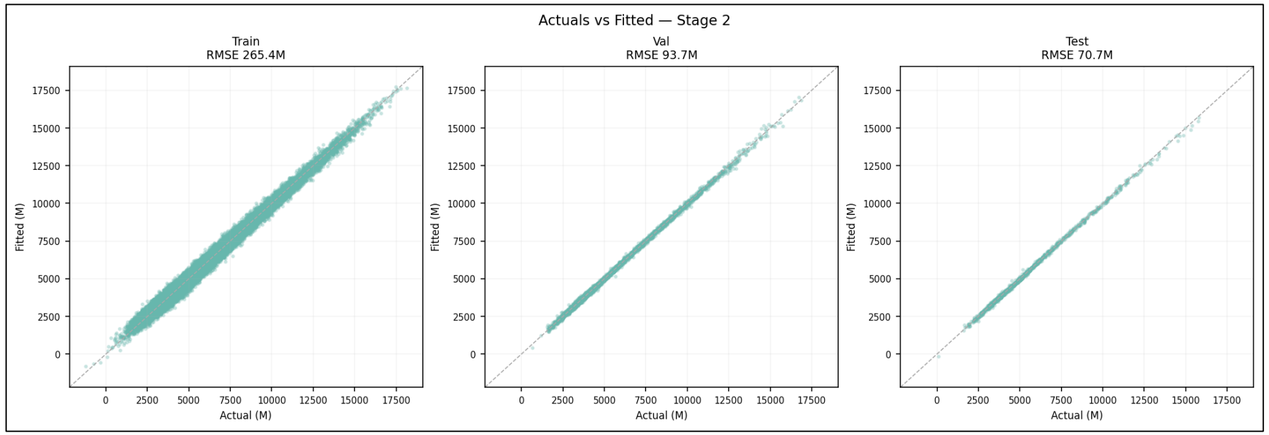


Human in the loop: Actuary to agree with the agent‘s proposal or caveat them or abort the process.

AI Agent Writing Report.

AI Agent generating a validation report covering all of the above.

Background #4: Production of Validation Report



1. Executive Summary

	Stage 1	Stage 2 (post-ML)
Val RMSE (95% CI)	163.0M	93.7M
Test RMSE (95% CI)	117.4M	70.7M
Margin to threshold	n/a	—

Recommendation: ACCEPT WITH CAVEATS

The Set1 proxy model calibration has been completed and the fit is recommended for acceptance with caveats. The two-stage model achieves a Stage 2 validation RMSE of £93.7M and an out-of-sample test RMSE of £70.7M, both within the £100M acceptance threshold. Hyperparameter optimisation was not run; XGBoost default parameters were used for the Stage 2 correction.

Stage 1 Legendre polynomial regression (order 4) produced a validation RMSE of £162.9M, materially exceeding the £100M threshold and independently triggering the ML correction criterion. Diagnostic assessment of Set1 (Section 3) identified five drivers with structural features requiring review:

- NYC_PC1 (x ≈ −0.90) — high-severity sharp kink (~90° gradient change); both detectors agree. Accepted at Gate A and treated as confirmed kink driver.
- Long_Imp (x ≈ −0.50) — high-severity sharp kink (~90° gradient change); both detectors agree. Accepted at Gate A and treated as confirmed kink driver.
- GAO_Takeup (x ≈ +0.75) — high-severity sharp kink, transitioning to a near-horizontal plateau; both detectors agree. Accepted at Gate A and treated as confirmed kink driver.
- FI_Vol (x ≈ −0.60) — medium-severity smooth local minimum; overridden at Gate A by the actuary as non-genuine; source investigation required.

A complete validation report being created, providing full audit trail and covering each stage of agentic work as well as each human decision.