Peer Review Report Manuscript Title: Insurance Pricing When Risks Are Artificially Generated: A Dynamic Control-Theoretic Framework for AI-Driven Hazards (Revised Manuscript) **Journal:** Annals of Actuarial Science (or equivalent high-quality actuarial periodical, e.g., ASTIN Bulletin, British Actuarial Journal) **Reviewer:** Grok (on behalf of the review team) **Date:** 21 February 2026

Recommendation: Accept with Minor Revisions

This revised manuscript represents a strong, publishable contribution to the emerging actuarial literature on pricing and capitalising risks generated by algorithmic systems. The authors have responded comprehensively and transparently to prior reviewer feedback, particularly by (i) adding an explicit bridge from the HJB formulation to an implementable premium (new Section 2.5), (ii) fully coupling the three model components (risk process R(t), Hawkes events N(t), claim severity), (iii) expanding calibration sources and sensitivity/robustness analysis, and (iv) clarifying the unit of analysis for concentration results. The result is a coherent, mathematically rigorous, and policy-relevant paper that advances the field beyond traditional AV or cyber models.

Summary of the Paper

The authors develop a unified stochastic framework for AI-driven hazards in which the risk level R(t) (Ornstein-Uhlenbeck with Hawkes-triggered jumps) endogenously influences both event intensity (state-dependent Hawkes) and claim severity (lognormal location shifted by R(t)). An HJB control problem yields the theoretical optimal premium; a transparent mean-variance + bounded tanh(·) approximation produces a regulator-friendly formula $\pi^*(R,t)$. Extensive Monte Carlo simulation of the coupled system demonstrates that uncoupled/traditional actuarial techniques understate VaR by 24–38% and CTE by 34–39%, while Gini coefficients rise from 0.25 to 0.58 over 10 years, indicating concentration rather than diversification. Practical elements include a calibrated parameter table referencing real data sources (CA DMV AV reports, Advisen/SAS OpRisk, Bessy-Roland et al.), tornado plots, systemic-correlation analysis, and historical analogues (Flash Crash, NotPetya, recent cloud outages).

Major Strengths

1. **Novelty and Relevance** (High): The core innovation—the *explicit functional coupling* of risk generation, self-excitation, and severity—directly addresses the “algorithmic liability” problem that standard Poisson/GLM or even basic Hawkes models miss. No comparable published work integrates stochastic control (HJB) with coupled endogenous dynamics for AI hazards (searches of actuarial and related literature confirm this; existing Hawkes applications in insurance are typically for cyber frequency only or uncoupled). The 2026 timing is ideal given accelerating AI deployment and regulatory focus (e.g., EU AI Act, US state AI bills).
2. **Methodological Rigour and Transparency**: The new Section 2.5 is exemplary: it openly acknowledges intractability of closed-form HJB solutions, derives the mean-variance loading from exponential utility + certainty equivalent, and justifies tanh as a bounded, smooth proxy to $\partial V/\partial R$. The unified data-generating process (Algorithm 2.1,

Figure 1) and stability conditions are clearly stated. Coupling is no longer "parallel" but mathematically consistent.

3. **Empirical Grounding and Robustness**: Calibration table (Table 1) with explicit sources is welcome. New sensitivity/robustness sections (2.7, tornado plot Figure 9, detailed Table in 3.10) and systemic-correlation analysis (4.5) strengthen credibility. Results are believable and actionable for capital setting and reinsurance.
4. **Actuarial and Policy Value**: Clear implications for regulatory capital, treaty design, and insurability monitoring. Discussion of unit of analysis (across AI systems/fleets/versions, not individual policies) and historical precedents avoids over-claiming while highlighting systemic risk.
5. **Presentation**: Well-structured (especially the expanded Response to Reviewers), readable for a mathematically inclined actuarial audience, with helpful figures and tables.

Minor Revisions and Suggestions (All Addressable in 4–6 Weeks)

Essential (must be done before acceptance)

1. **Observability/Estimation of Latent R(t)**: The premium formula conditions on $R(t)$, yet insurers will not observe the "risk level" of a black-box AI system in real time. Add a short subsection (e.g., 2.8.4 or in 4.7 Limitations) discussing practical proxies (telematics/disengagement rates for AVs, audit-log anomaly scores, model-card metrics, continuous monitoring APIs). Suggest a filtering approach (e.g., Kalman or particle filter) or credibility-weighted proxy. This is the single most frequent practical objection an actuarial reader will raise.
2. **Justification of Jump-Size Coupling**: $J_k = 0.1 \cdot \tanh(L/E[L|R])$ appears somewhat arbitrary. Either (a) calibrate the 0.1 scaling factor explicitly against empirical data (e.g., observed risk escalation post-incident in cyber/AV reports), or (b) generalise to a family $h(\cdot; \theta)$ and show robustness in sensitivity analysis. A one-sentence justification or reference would suffice.
3. **HJB Economic Interpretation**: Briefly clarify the surplus/process interpretation of the control problem (link to standard insurance-control literature, e.g., Schmidli 2008, or dividend problems). What exactly is the insurer controlling via $\pi(s)$? A short paragraph in 2.4 would help actuarial readers who are more familiar with classical risk processes.

Desirable (strongly recommended) 4. **Simulation Reproducibility**: Explicitly state number of paths (200 is mentioned), Δt, random-number generator/seed, and any tail-convergence diagnostics (e.g., QQ plots or bootstrap SE on VaR/CTE). Consider increasing paths to 1,000+ for 99%+ quantiles or note computational constraints.

5. **Literature Enhancements**: The review is solid but could note recent related work for completeness:
    - Swishchuk (2021) "Hawkes processes in insurance: Risk model..." (already in broader Hawkes context).
    - Ren (2025) on Hawkes with loss covariate for cyber pricing.

    - Liu (2026) "Insuring Algorithmic Operations" (liability pricing for algorithmic risk—complements but does not duplicate the dynamic control approach). These strengthen positioning without diluting originality.
6. **Minor Clarifications**
    - Abstract: Quantify the concentration result more precisely ("Gini rising to 0.58") for impact.
    - Premium formula (13)–(14): Define γAI and η explicitly in the text (or table).
    - Figure/Table cross-references: Ensure all (e.g., "Figure 10", "Table ??" in draft) are correctly numbered and captioned in final version.
    - Limitations (4.7): Expand slightly on strategic behaviour (moral hazard from AI providers optimising to minimise observed R(t); adverse selection across competing AI fleets).

Overall Assessment

This is now a mature, high-quality paper that an actuarial journal should be proud to publish. It moves the conversation from "AI will change insurance" to "here is a concrete, mathematically grounded way to price the new endogenous hazards." With the minor revisions above—primarily enhancing practicality and reproducibility—the manuscript will be ready for acceptance.

I am happy to review the revision if the editor wishes.

Confidential Comments to the Editor (not for authors) Excellent fit for AAS or similar; the control-theoretic angle and explicit response to reviewers demonstrate scholarly maturity. No ethical or originality concerns. The paper's length (28 pp.) is appropriate with the added sections. Priority: high, given topicality.

Please proceed with publication after the minor revisions outlined. This work will be cited.