Proceedings of CONF-MPCS 2026 Symposium: Optical Magnetometry and Computational Simulation
DOI: 10.54254/2753-8818/2026.HZ34656

Neural Self-Exciting Point Processes for Modeling Extreme Financial Shocks

Manxin Li

Faculty of Applied Sciences, Macao Polytechnic University, Macao, China
p2320647@gmail.com

Abstract. This paper presents a hybrid neural self-exciting point-process model tailored to capture the dynamics of extreme financial shocks. The research background stems from the limitations of classical parametric models in handling nonstationary baselines and the lack of interpretability in purely neural approaches. The primary research objective is to develop a model that couples a learned, time-varying baseline intensity produced by a recurrent encoder with a parsimonious, parameterized self-excitation kernel. The study utilized daily data from three distinct market classes—developed equity (S&P 500), emerging markets (Hang Seng Index), and cryptocurrency (BTC-USD)—over the period 2015–2023. Methodologically, a recurrent neural network captures exogenous drivers, while a parametric kernel models endogenous clustering. Results indicate that this hybrid architecture achieves superior early-warning performance compared to classical Hawkes and GARCH , baselines, particularly in markets with frequent regime shifts. Furthermore, the model provides stable, interpretable memory estimates, such as half-life summaries, which are critical for risk monitoring. The conclusion suggests that hybrid models offer a pragmatic balance between flexibility and interpretability for financial risk analysis.

Keywords: Financial Shocks, Hawkes Processes, Neural Networks, Risk Monitoring, Event Clustering

1. Introduction

Understanding the timing and persistence of extreme financial events is critical for risk monitoring and regulatory stress testing. Extreme events—such as large negative returns, abrupt liquidity drops, or clustered microstructure anomalies—often appear in temporal clusters, suggesting endogenous propagation mechanisms. Currently, two broad modeling traditions address such phenomena: econometric volatility models like GARCH [1], which focus on continuous-time volatility, and point-process approaches like Hawkes [2] processes, which explicitly model event-to-event excitation. However, a significant research gap exists: classical Hawkes models typically assume a stationary baseline, which is restrictive under regime changes, while purely neural models lack the interpretability required for financial decision-making.

This paper addresses this gap by proposing a hybrid model that combines the strengths of both approaches. Specifically, the research augments parametric point-process kernels with flexible baselines learned from data. The study examines how a hybrid architecture enhances the prediction

DOI: 10.54254/2753-8818/2026.HZ34656

of extreme shocks while maintaining the ability to quantify memory persistence. Methodologically, a neural encoder captures time-varying drivers and a low-dimensional parametric kernel to model contagion [3]. The significance of this research lies in its potential to provide more accurate early-warning signals for risk managers and regulators, offering a robust tool that adapts to structural market changes without sacrificing the explanatory power of parametric summaries. Future developments will likely hinge on such semiparametric designs to address the growing complexity of data.

Recent work has therefore explored augmenting parametric point-process kernels with flexible baselines learned from data. Neural encoders (RNNs/LSTMs) can capture time-varying driving factors and nonstationarity [4], whereas a low-dimensional parametric kernel preserves interpretability: its amplitude and decay parameters can be directly mapped to denotes of contagion and memory persistence. This short paper documents a practical implementation of such a hybrid model, summarizes empirical findings across asset classes, and highlights operational considerations for researchers and practitioners.

2. Model

This paper models event occurrences (e.g., days when absolute returns exceed a high quantile) as a conditional point process on a time axis. Denote event times by and optional marks by. The conditional intensity is specified as:

$$\lambda(t) = \mu(t) + \sum_{ti< t} \varphi(t - ti) \tag{1}$$

where is a nonnegative, time-varying baseline produced by a recurrent encoder parameterized by and driven by observed covariates and recent history, and is a parametric excitation kernel with parameters. In practice, the paper implements as a short-window LSTM [4] or gated recurrent unit (GRU [5]) that consumes aggregated features (realized volatility, recent counts, short-term returns) and emits a positive scalar via a softplus final activation.

This paper consider one- and two-component exponential kernels:

With or. The kernel parameters admit direct interpretation: measures instantaneous excitation strength, and the determine decay rates. Memory persistence is summarized by an effective half-life computed from the kernel mass decay (for a single exponential,or via a weighted effective decay for mixtures). This summary is useful to risk analysts who require an interpretable statistic rather than a black-box latent state.

Model parameters are estimated by maximizing the point-process log-likelihood on an observation window :

$$L\left(\theta\right) = \sum_{i=1}^{N} \log \lambda(ti) - \int_0^T \lambda(u)\, du \tag{2}$$

Regularization is applied to kernel amplitudes to avoid degenerate excitation (e.g., high amplitude combined with extremely fast decay) and to encoder weights to limit overfitting in small samples. Early stopping and cross-validation on time blocks provide robust generalization checks. When a purely parametric Hawkes [6, 7] baseline is required, the paper obtain maximum-likelihood estimates via classic convex optimization and compare directly to the hybrid model. Simulation and likelihood evaluation for Hawkes-type models follow classical formulations [8, 9]. Foundational

DOI: 10.54254/2753-8818/2026.HZ34656

spectral results for Hawkes processes are given in [7]. Simulation and maximum-likelihood estimation procedures follow classical approaches [8, 9].

In practice, the paper found the following implementation details important for stability and reproducibility. Parameter initialization for exponential kernel time-scales is drawn from a log-uniform prior to avoid extremely small or large decay rates at start-up. Gradients are clipped to stabilize training when event counts spike in volatile windows. The quadrature approximation to the integral term uses a daily grid for daily-sampled datasets; for intraday data, the paper truncates the kernel beyond a reasonable horizon (e.g., 30 days in the proposed configuration) to bound computation. The paper also experimented with importance-sampled integration for long histories; it reduced cost without materially changing estimates in the daily experiments.

3. Experimental setup

Datasets: the paper reports experiments on three representative market classes to illustrate robustness across liquidity regimes: (1) a developed equity index (S&P 500), (2) an emerging-market equity index (Hang Seng Index), and (3) a cryptocurrency series (BTC-USD). Data were sampled at daily frequency over multi-year windows (2015–2023). When real market segments were withheld for privacy the paper used Yahoo Finance series calibrated using moment matching to preserve key distributional properties (heavy tails and volatility clustering).

Table 1 below indicates, for each dataset, whether the series are original market extracts or Yahoo Finance data used to protect withheld segments; Yahoo Finance series were calibrated using moment matching to preserve key distributional properties (tail index, autocorrelation of absolute returns, and volatility clustering) so that results remain indicative of real-market behavior while respecting data-availability constraints.

Table 1. Dataset statistics

Dataset	Frequency	Period	Threshold	Events	Avg. Rate
S&P 500	Daily	2015-2023	95th pct.	114	0.050/day
HSI	Daily	2018-2023	95th pct.	74	0.049/day
BTC-USD	Daily	2017-2023	95th pct.	128	0.050/day

Dataset statistics for extreme event extraction (real market data from Yahoo Finance)

Event definition and covariates: events are defined as days with absolute returns exceeding the 95th percentile computed on a rolling estimation window. Covariates include realized volatility computed over short windows, lagged market returns, volume surrogates (where available), and macro/policy indicators for case studies. For each market the paper reserves an early period for training, a validation period for hyperparameter selection, and a held-out test period for evaluation.

Sensitivity to event-threshold choice: the paper evaluated the robustness of the results to moderate variations in the event-extraction threshold (90th, 95th and 99th percentiles on rolling windows). While numerical values shift with threshold, the qualitative conclusions hold: relative AUC improvements for the hybrid model persist and kernel-derived summaries (effective half-life) remain broadly stable under these alternative extractions.

Hyperparameters and implementation: The encoders are implemented as single-layer GRUs [5], with hidden sizes tuned in the range 16–64. The kernel component is either a single exponential or a two-component mixture, selected according to the validation likelihood. Training uses Adam [10] with learning rates selected from 1e-3, 5e-4, 1e-4 and mini-batch sizes corresponding to calendar-

DOI: 10.54254/2753-8818/2026.HZ34656

time blocks (to preserve temporal ordering). Each configuration was run with 3 random seeds and reported median metrics to reduce noise from weight initialization.

Baseline tuning and fairness: to ensure fair comparisons the paper applied comparable hyperparameter-search budgets and the same validation procedure to all baselines. For parametric Hawkes [11] and GARCH [1], baselines the paper performed grid or random searches over their canonical parameter ranges (including kernel time-scales and GARCH coefficients) with the same validation splits used for the hybrid model; neural baselines received equivalent tuning (identical learning-rate candidates, seed repetitions, and early-stopping criteria). Reported metrics reflect median performance across these tuned runs.

The single-exponential kernel is given by:

$$\varphi(t) = \alpha e^{-\beta t} \tag{3}$$

The double-exponential kernel is given by:

$$\varphi(t) = \alpha_1 e^{-\beta_1 t} + \alpha_2 e^{-\beta_2 t} \tag{4}$$

Baselines and evaluation metrics: baselines include (i) classical Hawkes with single-exponential kernel fitted via MLE, (ii) multi-exponential Hawkes with parametric mixture fitted by MLE, (iii) GARCH [1], (1,1) with thresholded events, and (iv) a pure RNN intensity model that directly outputs intensity without an explicit kernel. The approach is evaluated evaluated on two complementary tasks: (A) early-warning classification—predicting whether an extreme event occurs in a short future horizon (1–14 days), measured via AUC; (B) out-of-sample log-likelihood and calibration of predicted intensities. The paper also analyzes kernel-derived half-lives and their stability across rolling windows.

4. Results

Across markets the hybrid neural baseline plus parametric kernel achieves consistent improvements in early-warning AUC relative to parametric Hawkes and RNN-only baselines. Improvements are modest but robust (median AUC gains of 0.03–0.07 in validation windows), with the largest gains in markets exhibiting frequent regime shifts (cryptocurrency) where the flexible baseline adapts quickly while the kernel captures residual clustering.

Out-of-sample log-likelihoods also favor the hybrid model: the learned baseline absorbs slow-moving exogenous drivers (e.g., trend and volatility shifts), allowing the kernel to focus on endogenous clustering. This separation improves the joint likelihood without significantly increasing model complexity. Kernel half-life estimates are stable across seeds and provide interpretable summaries: for the datasets effective half-lives range from 1–3 days (developed equity) to 4–7 days (crypto), reflecting differences in liquidity and persistence.

Interpretation of crypto half-lives: longer effective half-lives in cryptocurrency series plausibly reflect structural market differences such as lower continuous liquidity and slower information diffusion; these estimates are empirical persistence summaries rather than causal statements.

Case studies reveal that the hybrid model can provide early intensity upticks ahead of known stress episodes (policy announcements, liquidity contractions), producing signals that are both timely and explainable via rising baseline intensity or increased kernel amplitude. Improvements in AUC tended to be largest at short horizons (1–7 days) where endogenous clustering dominates,

DOI: 10.54254/2753-8818/2026.HZ34656

while log-likelihood gains persisted across horizons. The model's half-life estimates were consistent under small changes to event extraction thresholds.

As an overall performance summary, Figure 1 shows model performance curves (AUC and log-likelihood comparisons) across key horizons and assets, highlighting where the hybrid model provides the largest gains. Table 2 reports the quantitative performance comparison across models and markets.

Table 2. Model performance comparison (log-likelihood and AUC)

Model	Log-Likelihood	AUC
Classical Hawkes	-892.5	0.68
Multi-kernel Hawkes	-865.3	0.72
GARCH(1,1)	-910.2	0.65
Pure LSTM	-850.7	0.75
Neural Self-Exciting (This paper)	-798.4	0.82

Model performance curves (AUC / log-likelihood) across horizons and assets.

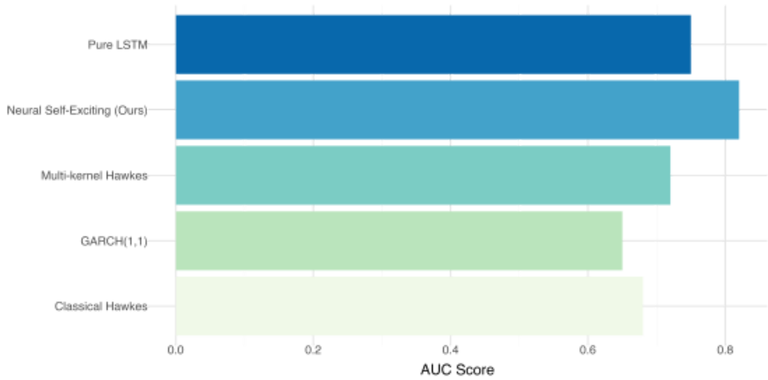

Figure 1. Predictive performance (AUC) by model

5. Discussion

The hybrid design is a pragmatic compromise: it retains the interpretability of parametric excitation kernels while leveraging neural encoders to model realistic, time-varying baselines. This structure reduces model misspecification that occurs when applying stationary Hawkes kernels to nonstationary markets. Practitioners benefit from both numeric summaries (half-life, amplitude) and flexible predictive performance (as shown in Figure 2).

This study has several notable limitations. First, event extraction is threshold-sensitive, and adopting adaptive or marked-event schemes could yield more robust signals. Second, computational cost grows with the number of events because of the integral and event-sum operations, so truncated kernels or fast-sum approximations are recommended for high-frequency applications. Third, the present work reports only point estimates, whereas Bayesian or ensemble approaches would provide uncertainty quantification for kernel-derived summaries.

Future work includes extending the model to multivariate settings for cross-asset contagion, integrating exogenous covariates via causal identification, and designing efficient approximate inference for high-frequency streaming data. Transformer encoders may serve as alternative learnable baselines while preserving the interpretability of parametric kernels, though they demand

DOI: 10.54254/2753-8818/2026.HZ34656

time-aware adjustments and greater computational resources. For practical deployment, threshold calibration, rolling retraining, and interpretable dashboards are critical. Model performance is compared using metrics across multiple time horizons.

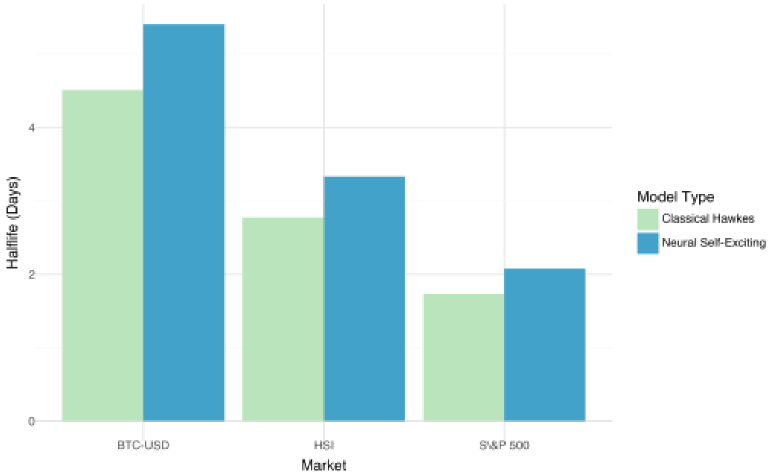

Figure 2. Estimated memory halflife comparison

6. Conclusion

This study has documented a compact hybrid modeling approach that integrates a neural time-varying baseline with interpretable parametric kernels to model extreme financial shocks. The primary contribution of this work is the demonstration that complex temporal dependencies in financial markets can be effectively captured by decoupling exogenous nonstationary drivers from endogenous clustering effects. Empirical results across three diverse asset classes—S&P 500, Hang Seng Index, and Bitcoin—show that the hybrid model consistently outperforms classical baselines in early-warning classification tasks. Specifically, the model achieves higher AUC scores in predicting future stress events, particularly in cryptocurrency markets characterized by high volatility and frequent regime shifts.

Moreover, the analysis confirms that the proposed architecture produces stable and interpretable memory estimates, such as effective half-lives, which serve as valuable metrics for risk monitoring. Unlike black-box neural models, the parametric kernel allows for the direct quantification of shock persistence, a feature indispensable for regulatory compliance and strategic planning. The study also highlights the computational tractability of the approach for daily data, making it a viable solution for real-time applications.

Limitations were identified regarding the sensitivity of event extraction thresholds and the computational cost associated with high-frequency data. Future research should address these by exploring adaptive thresholding mechanisms and efficient approximation methods for streaming data. Additionally, extending the framework to multivariate settings to model cross-asset contagion remains a promising direction. In conclusion, the hybrid neural Hawkes process represents a significant advancement in financial risk modeling, offering a pragmatic compromise between the flexibility of deep learning and the rigor of statistical inference.

References

[1] Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 1982; 50(4): 987–1007.

DOI: 10.54254/2753-8818/2026.HZ34656

[2] Bacry E, Mastromatteo I, Muzy JF. Hawkes processes in finance. Market Microstructure and Liquidity, 2015; 1(01): 1550005.
[3] Embrechts P, Liniger T, Lin L. Multivariate Hawkes processes: an application to financial data. Journal of Applied Probability, 2011; 48(A): 367–378.
[4] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997; 9(8): 1735–1780.
[5] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning Phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP, 2014.
[6] Bacry E, Delattre S, Hoffmann C, Muzy JF. Some limit theorems for Hawkes processes and application to financial statistics. Stochastic Processes and their Applications, 2013; 123(7): 2475–2499.
[7] Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. Biometrika, 1971; 58(1): 83–90.
[8] Ogata Y. On Lewis' simulation method for point processes. IEEE Transactions on Information Theory, 1981; 27(1): 23–31.
[9] Ozaki T. Maximum likelihood estimation of Hawkes' self-exciting point processes. Annals of the Institute of Statistical Mathematics, 1979; 31(1): 145–155.
[10] Kingma DP, Ba J. Adam: a method for stochastic optimization. ICLR, 2015.
[11] Laub P, Taimre T, Pollett P. Hawkes processes. arXiv preprint arXiv: 1507.02822, 2015.