

UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Intraday Jump Information and Realized-Volatility Forecasts

Estimation, testing and matched evidence from Hawkes processes

Diego Alejandro Astaburuaga Corveleyn

Como requisito para optar al título profesional de
Ingeniero Civil Matemático
Universidad Técnica Federico Santa María

Director de Tesis: Dr. Werner Kristjanpoller
Profesor Correferente: Dr. Francisco Cuevas

Octubre 2026

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Intraday Jump Information and Realized-Volatility Forecasts: Estimation, testing and matched evidence from Hawkes processes

Nombre del candidato(a): Diego Alejandro Astaburuaga Corveleyn

Carrera / Grado: Ingeniería Civil Matemática / Ingeniero Civil Matemático

Campus: Casa Central Valparaíso **Departamento:** Matemática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Werner Kristjanpoller, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

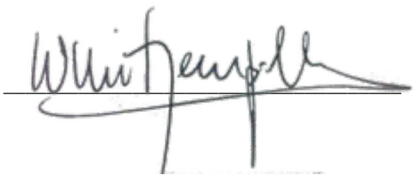
☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 02/11/2026 **Firma:** 

Estudiante o Candidato(a):

Fecha: 02/11/2026 **Firma:** 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Agradecimientos

Dedico esta memoria y los frutos de mi formación a mis amados padres, Sandra Corveleyn Saavedra y Claudio Astaburuaga Contreras, por su sacrificio y apoyo incondicional, y por no solo darme una carrera profesional, sino también permitirme crecer y buscar mi propio camino personal durante todo el proceso.

A mis amigos y compañeros de universidad, por acompañarme y ayudarme a crecer durante estos años. Al Departamento de Matemática y sus profesores, por guiarme en el sumamente interesante mundo de las matemáticas, y a la Universidad Técnica Federico Santa María, por permitirme estudiar y brindarme vastas oportunidades para seguir formándome y crecer profesionalmente.

A mi profesor guía, Werner Kristjanpoller, por su orientación y enseñanza en finanzas cuantitativas, y a mi profesor correferente, Francisco Cuevas, por su apoyo desde el Departamento de Matemática. A Inria Chile, a Inria Nancy – Grand Est y al profesor Radu Stoica, por permitirme ampliar mis horizontes hacia la investigación internacional. A Asier Uriarte, a María Fernanda Pérez y al equipo de la mesa de dinero del Banco Bci, por abrirme las puertas de la industria financiera chilena y mostrarme sus desafíos.

Finalmente, cierro esta etapa con gratitud, mirando con expectativa e interés el camino profesional que comienza.

Resumen

La magnitud de los saltos pasados aporta poco al pronóstico de la volatilidad realizada, pero el momento en que ocurren podría aportar más. Esta tesis evalúa si la información sobre el momento y la agrupación de los saltos intradiarios mejora los pronósticos a un día de la volatilidad realizada de Bitcoin. Para ello se ajustan procesos de Hawkes a los saltos detectados en precios de cinco minutos de Binance y se incorpora el número esperado de saltos del día siguiente a un modelo HAR con componentes continuo y de saltos, comparando todos los pronósticos sobre los mismos orígenes.

La intensidad al cierre del día reduce el error cuadrático del logaritmo en 1,16 % y el error relativo en 4,67 %, a diferencia de las intensidades diaria e integrada. Un segundo componente exponencial mejora el ajuste de las llegadas, pero no el pronóstico. Un conteo descontado con memoria calibrada rinde igual, de modo que estimar el proceso de Hawkes no es necesario. La varianza reciente de todos los retornos reduce el error en 12–13 %, y sumarle la intensidad de saltos no agrega una mejora significativa. Las ganancias no se mantienen a siete ni treinta días, en Ethereum y Litecoin se limitan al error relativo, y no se traducen en ventajas económicas.

El momento de los saltos es útil frente a HAR-CJ, pero la varianza reciente de todos los retornos recoge la mayor parte de esa información y debe incluirse en el modelo de referencia. Se proponen como líneas futuras modelos de Hawkes discretos, con núcleos de ley de potencia, multivariados, no lineales y de picos sobre umbral.

Palabras clave: volatilidad realizada; saltos; procesos de Hawkes; evaluación de pronósticos; Bitcoin.

Abstract

Past jump variation adds little to realized-volatility forecasts, yet the timing of jumps may carry information that their size does not. This thesis evaluates whether information about the timing and clustering of intraday jumps improves one-day Bitcoin realized-volatility forecasts. Hawkes processes, in which an event temporarily raises the rate of later events, are fitted to five-minute jumps detected on Binance data, and the conditional expected jump count of the next day enters a heterogeneous autoregressive (HAR) model with continuous and jump components. All forecasts are compared on matched rolling origins.

The closing intraday intensity lowers the benchmark's squared log error by 1.16% and its relative squared error by 4.67%, while the daily and integrated intensities do not improve it. A second exponential component improves the fit to jump arrivals at every refit without a significant forecast gain. A decayed jump count with forecast-tuned memory is equally competitive, so Hawkes estimation is not necessary for the gain. Recent variance measured from all returns reduces the squared log error by about 12–13%, and adding the jump intensity to it gives no significant further improvement under any loss. The gains do not extend to seven- or thirty-day horizons, Ethereum and Litecoin share only the relative-error gains, and the allocation and Value-at-Risk exercises show no economic advantage.

Jump timing is useful relative to HAR-CJ, but most of its information is carried by recent all-return variance, which should therefore be part of the benchmark before a fitted jump process is added. Discrete-time, power-law, multivariate, nonlinear and peaks-over-threshold Hawkes specifications are proposed as future research.

Keywords: realized volatility; jumps; Hawkes processes; forecast evaluation; Bitcoin.

Contents

Agradecimientos ii

Resumen iii

Abstract iv

Notation and Abbreviations xiii

1 Introduction 1

1.1 Research Context and Practical Motivation 1
1.2 Realized Volatility Forecasting 2
1.3 Cryptocurrency Market and Bitcoin Volatility 4
1.4 Hawkes Process Applications 5
1.5 Limits in the Literature 5
1.6 Investigation Statements 6
1.7 Main Contributions and Thesis Outline 7

2 Theory and Methods 9

2.1 Realized Volatility and Jump Measures 9
2.1.1 Quadratic variation and realized measures 9
2.1.2 Jump detection 11
2.2 Temporal Point Processes 13
2.2.1 Conditional intensity and Poisson Process 13
2.2.2 The Hawkes process 14
2.2.3 Richer exponential self-exciting 15
2.2.4 Maximum likelihood estimation 16
2.2.5 Simulation 18
2.2.6 Likelihood ratio test 19
2.3 From a Fitted Process to a Daily Regressor 20
2.3.1 The expected count over a forecast window 20

2.3.2 The regressors . . . 21
2.4 Forecasting Realized Volatility . . . 22
2.4.1 The HAR family . . . 22
2.4.2 HAR-H: a jump-arrival summary as the additional regressor 23
2.5 Forecast Evaluation . . . 24
2.5.1 Out-of-sample schemes . . . 24
2.5.2 Loss functions . . . 24
2.5.3 Pairwise comparison . . . 25
2.5.4 Model confidence set . . . 26
2.5.5 Horizons longer than one day . . . 27
2.5.6 Economic significance . . . 27

3 Empirical Application 29
3.1 Data Setup and In-Sample Analysis . . . 29
3.1.1 Data Sources and Return Construction . . . 29
3.1.2 Realized Measures and Jump Definitions . . . 30
3.1.3 In-Sample Analysis . . . 31
3.2 Out-of-Sample Forecasting . . . 34
3.2.1 Forecast Design and Evaluation . . . 34
3.2.2 Comparison of Jump-Arrival Regressors . . . 35
3.2.3 Ethereum and Litecoin . . . 36
3.3 Fit and Forecast . . . 37
3.3.1 Self-Excitation against a Poisson Process . . . 37
3.3.2 Richer Processes and Forecast Accuracy . . . 39
3.4 The All-Return Controls . . . 40
3.4.1 The Two Controls . . . 40
3.4.2 Three Complementary Tests . . . 42
3.4.3 The Last Hours of the Day . . . 44
3.4.4 The Hour of the Jump . . . 45
3.5 Robustness . . . 45
3.5.1 Jump Tests, Significance Levels and Sampling Frequency 47
3.5.2 The Controls on Denser Event Sets . . . 48
3.5.3 HAR Cascade and Refit Schedule . . . 48
3.5.4 Dependence and Multiplicity . . . 48
3.5.5 The Day Boundary and the HAR Specification . . . 50
3.5.6 Horizons of Seven and Thirty Days . . . 51
3.6 Economic Significance Exercise . . . 52

4 Conclusion **54**
4.1 Main Results and Scope 54
4.2 Interpretation 55
4.3 Limitations and Boundaries 56
4.4 Directions for Future Research 57

A Derivations and Additional Definitions **59**
A.1 Gradient of the Negative Log-Likelihood 59
A.2 Derivation of the One-Exponential Expected Count 60
A.3 Expected Count of the Power-Marked Bi-Exponential Process . 60
A.4 The HAR-RS and LHAR-RV Models 63
A.5 Economic Definitions 64

B Simulation and Numerical Validation **66**
B.1 Design 66
B.2 The Simulator 68
B.3 The Count of a Stationary Process 69
B.4 Time Rescaling at the True Parameters 72
B.5 The Estimator 73
B.6 The Test Statistics at Estimated Parameters 75
B.7 The Expected Count from a Given State 79
B.8 Count Verification and Plug-In Uncertainty 80
B.9 The Likelihood Ratio 82
B.10 Empirical Goodness of Fit on Bitcoin 84

C The S&P 500 Application **85**
C.1 Data and Session Clock 85
C.2 Replication of the Literature on the S&P 500 86
C.3 The Model Ladders 89
C.4 Ten-Minute Clements–Liao Configuration 91
C.5 Five-Minute Qu Configuration 92
C.6 All-Return Controls 93
C.7 Sampling Sensitivity 93

D Supplementary Cryptocurrency Evidence **95**
D.1 Alternative Full-Sample Regressions 95
D.2 Regressor and Confidence-Set Sensitivity 96
D.3 Richer Processes and State Summaries 96

D.4 Additional Control and Boundary Comparisons 99

E Data and Computational Reproducibility **102**

References **104**

List of Tables

3.1 The stored price series and their coverage 30
3.2 Distribution of the daily realized measures: Bitcoin 31
3.3 Jump days and intraday jump events by asset and detector . . . 32
3.4 Full-sample Hawkes fits of the model ladder 33
3.5 Full-sample fit of HAR-CJ and its extensions: Bitcoin 33
3.6 Forecast samples by asset and refit schedule 34
3.7 Primary intraday comparison: Bitcoin 35
3.8 Main-configuration model confidence set p-values with statistic-specific retention prevalence. 36
3.9 Primary intraday comparison: Ethereum and Litecoin 38
3.10 Self-excitation against a Poisson process on rolling windows: Bitcoin 38
3.11 Fit against forecast: richer Hawkes processes and their regressors 39
3.12 The all-return controls and the jump regressor: Bitcoin 41
3.13 The all-return controls and the jump regressor: Ethereum and Litecoin 41
3.14 Whether the jump regressor adds to an all-return control 42
3.15 Forecast encompassing of the jump regressor by the controls . . 43
3.16 The memory of the all-return control 44
3.17 The gain of the jump regressor by the time since the last jump . 46
3.18 Moving each jump to another time of its day: Bitcoin 46
3.19 Sampling frequency and event definition: the intraday regressors against HAR-CJ 47
3.20 Sensitivity to the significance levels of the jump tests 48
3.21 The all-return controls on the sparse and the dense event sets: Bitcoin 49
3.22 The HAR cascade: 1/5/22 against 1/7/30 49
3.23 Serial dependence and multiplicity in the loss comparisons: Bitcoin 50
3.24 The day boundary and the backbone: Bitcoin 51

3.25 Forecasts of the mean realized variance over seven and thirty days: Bitcoin . . . 51
3.26 Economic significance exercise: a portfolio under a risk target, Bitcoin . . . 52
3.27 Economic significance exercise: one-day Value-at-Risk forecasts, Bitcoin . . . 53

B.1 Mean, variance and normality of the simulated count against their theoretical values . . . 70
B.2 Identities of the likelihood: largest relative deviation across simulated paths . . . 71
B.3 Rejection rates of the exponential and independence tests on durations rescaled at the true parameters . . . 73
B.4 Recovery of the parameters by maximum likelihood on simulated paths . . . 75
B.5 Plug-in goodness-of-fit statistics against their known-parameter critical values . . . 76
B.6 Rejection rates of the bootstrap goodness-of-fit test under a true model, in continuous time and on a time grid . . . 77
B.7 A true model observed on a time grid: events lost, fitted parameters and the Kolmogorov–Smirnov statistic . . . 79
B.8 The expected count against the mean of simulated counts, from six states at the origin . . . 80
B.9 The expected count against the mean of simulated counts: four configurations . . . 80
B.10 Bias and dispersion of the expected count evaluated at estimated parameters: four configurations . . . 81
B.11 The likelihood ratio of the squared-return mark on rolling windows: asymptotic, bootstrap and grid-aware bootstrap decisions . . . 83

C.1 Designs underlying the main forecast comparisons . . . 85
C.2 The S&P 500 series and its event sets . . . 86
C.3 Full-sample maximum likelihood estimates of the unmarked exponential Hawkes processes: S&P 500 . . . 87
C.4 Full-sample estimates of the HAR-CJ regressions on the S&P 500 88
C.5 Primary intraday comparison: S&P 500 . . . 89
C.6 Ten-minute literature-based ladder on the S&P 500 . . . 91
C.7 Five-minute daily intensity ladder on the S&P 500 . . . 92

C.8 The all-return control on the S&P 500 93

D.1 Full-sample estimates of the HAR-RV regressions on Bitcoin . . 95
D.2 The regressors of the literature against HAR-CJ: Bitcoin, by event set and sampling frequency 96
D.3 Model confidence sets across configurations: Bitcoin 97
D.4 The two-exponential process as a regressor, in four forms: Bitcoin 97
D.5 Full-sample Hawkes fits on every event set 100
D.6 The two controls with the events of the ABD test: Bitcoin . . . 101
D.7 The all-return control under a change of the day boundary . . . 101

List of Figures

3.1 Loss of the decayed count as a function of its half-life. The count enters HAR-CJ at a half-life held fixed over the whole forecast period; each point is the percentage change in MSE against HAR-CJ under the primary daily-refit design. Dashed line: the median half-life of the daily likelihood fits of the intraday process. Dotted line: the median half-life of the daily tuning of F, drawn where it falls inside the range of the axis. The vertical axis is cut at 0.6; the five-minute point of Bitcoin and the five- and ten-minute points of Ethereum lie above it. 37

3.2 Where the gain of an all-return regressor sits. Left: EWRV enters HAR-CJ at a half-life held fixed over the whole forecast period, from five minutes to 22 days. Right: the realized variance of a one-hour window that ends the stated number of hours before the close; zero is RV60. Each point is the percentage change in MSE against HAR-CJ, main design. Windows ending two and eight hours before the close contain a day with no price change and are not shown. 45

D.1 Cumulative loss gains against HAR-CJ under the primary daily-refit design. Each curve is $100 \sum_{s \le t}(L_{0,s} - L_{m,s}) / \sum_{s=1}^{n} L_{0,s}$; its endpoint equals the percentage gain in the corresponding overall table. Positive values favor the extension. MSE uses log RV; HRMSE uses the squared relative level error. The curves describe when loss differences accrue; they are not sequential significance tests. 98

Notation and Abbreviations

Realized measures

$r_{t,i}$, M	intraday return i of day t, in percent; number of intraday returns of a day
RV_t, BPV_t	realized variance and bipower variation (Section 2.1.1)
C_t, J_t	continuous and jump components (Section 2.1.1)
$MedRV_t$	median realized variance ((2.18))
RS_t^+, RS_t^-, r_t^-	realized semivariances and negative part of the daily return (Definition A.2)
$X_t^{(k)}$	average of the daily series X over the k days ending at t
$\mathcal{J}_t$	Intersection event set of day t (Section 2.1.2)

Point processes

t_q, $N(A)$, n	event times, number of events in the set A, number of events in the window
$\mathcal{H}_t$	history of the process up to and including t
$\lambda^*(t)$, Λ^*	conditional intensity and compensator (Section 2.2.1)
μ, α, β	baseline, impact and decay of the exponential kernel (Section 2.2.3)
ξ_q, δ	mark of event q and exponent of its impact
η, λ_∞	branching ratio and stationary intensity (Section 2.2.3)
β_*	effective decay, $\beta(1-\eta)$ (Section 2.3.1)
$\log 2/\beta$	half-life of the excitation

Estimation and testing

θ, Θ, $\hat{\theta}$	parameter vector, admissible box and maximum likelihood estimate ((2.2.4))
$\ell(\theta)$, $\mathrm{NLL}_g(\theta)$	log-likelihood and negative log-likelihood of the ground process ((2.23))

T	length of the observation window, in days
τ_q	rescaled durations (Proposition B.4)
LR	likelihood ratio statistic (Section 2.2.6)

Regressors

λ_t	intensity at the close of day t, $\lambda^*(t^+)$
$\mathbb{E}\Lambda_t(h)$	expected number of events over the h days after the close of day t (Section 2.3.1)
EI_t, LCI_t, SI_t, DI_t	intensity at the close (in levels, linearly equivalent to the expected count of the next day), logarithm of that expected count, integrated intensity of the day, daily intensity (Section 2.3.2)
$F_t(h_F)$, DF_t	decayed count of the events with half-life h_F, and the same count of the jump days
$EIbi_t$, $EIbi_t^+$, $EIbi_t^-$	expected count of the two-exponential process and its fast and slow components (Definition A.1)
$EWRV_t(h_E)$	exponentially weighted realized variance with half-life h_E (Section 2.3.2)
$RV60_t$	realized variance of the last hour of day t (Section 2.3.2)

Forecast evaluation

y_t, $\widehat{y}_t$, $\widehat{V}_t$	$\log RV_t$, its forecast, and the forecast of the level
d_t, DM	loss differential and Diebold–Mariano statistic ((2.38))
w_t, σ^*, c	portfolio weight, annual risk target and proportional cost ((A.16))
R_t^p, W_t	net return of the portfolio and its wealth ((A.17))
CE, γ	certainty-equivalent return and risk aversion (Definition A.3)
$\mathrm{VaR}_{q,t}$	Value-at-Risk at level q (Definition A.4)
z_q, ρ_q	quantile of the standardized return and quantile loss ((A.20))

Abbreviations

HAR	heterogeneous autoregression; HAR-RV on realized variance
HAR-CJ	HAR with continuous and jump cascades, the benchmark
HAR-RS, LHAR-RV	HAR with signed semivariances; HAR with leverage terms
HAR-H	HAR-CJ with a jump-arrival regressor
BNS, ABD, LM, sBNS	the jump tests of Barndorff-Nielsen and Shephard, of Andersen, Bollerslev and Dobrev, of Lee and Mykland, and the sequential BNS test

MLE, L-BFGS-B	maximum likelihood estimation; the limited-memory quasi-Newton method with bounds
KS, CvM	Kolmogorov–Smirnov and Cramér–von Mises statistics
DM–HLN	Diebold–Mariano test with the Harvey–Leybourne–Newbold correction
MSE, MAE	squared and absolute error of the logarithm
QLIKE, HRMSE	quasi-likelihood loss and squared relative error of the level
BTC, ETH, LTC, SPX	Bitcoin, Ethereum, Litecoin and the S&P 500 index
UTC	Coordinated Universal Time

Chapter 1

Introduction

1.1 Research Context and Practical Motivation

Volatility forecasting is important for risk management, portfolio allocation, and derivative pricing. For example, it can help determine an asset's contribution to portfolio variance and improve Value-at-Risk forecasts, helping monitor potential losses in extreme scenarios (Markowitz, 1952; Jorion, 2001). Forecasts therefore affect a range of financial decisions.

With the development of intraday price data, volatility can be measured using returns observed within a day rather than inferred only from daily returns (Andersen et al., 2001b; Andersen et al., 2003). Daily quadratic variation has consequently become an important measure of volatility. Advances in the theory have also made it possible to disentangle its continuous and jump components, which realized measures estimate separately (Barndorff-Nielsen and Shephard, 2004; Andersen et al., 2007a). Meanwhile, the heterogeneous autoregression (HAR) model of Corsi (2009) has become a strong benchmark for forecasting these measures. Its extensions use this decomposition to examine how past continuous and jump variation enter the model separately.

That decomposition exposes a puzzle (Qu et al., 2024). Most persistence in realized variance belongs to the continuous component, while past jump variation adds little once the continuous component is included (Andersen et al., 2007a; Corsi et al., 2010). Many studies have examined whether jump dynamics improve volatility forecasts, yet evidence of their predictive value remains limited across equity, fixed-income, foreign-exchange, and energy markets (Andersen et al., 2007a; Qu et al., 2024). For example, Andersen et al. (2007a) find that jump sizes are neither persistent nor predictable and do not significantly improve volatility

forecasts. Corsi et al. (2010) suggest that this weak predictive contribution may partly reflect small-sample bias in bipower variation and market microstructure effects. However, the puzzle remains unresolved: jumps are salient large price changes, yet explicitly modeling them has often yielded little improvement in forecasting accuracy.

The timing of jumps offers a different summary from their accumulated size (Clements and Liao, 2017). A Hawkes process (Hawkes, 1971) raises its conditional arrival intensity after an event and lets that increase decay, so fitted intensity can summarize recent clustering in jump arrivals. In variance forecasting, Clements and Liao (2017) show that point-process jump intensity adds information beyond past jump measures, while later HAR-type applications bring jump intensity into volatility forecasts for crude-oil futures (J. Wang et al., 2021; Qu et al., 2024). The forecasting question here is whether this timing-based summary contributes information beyond the usual realized-variance regressors, and, in a richer specification, beyond recent variation measured from every return.

Bitcoin (Nakamoto, 2008) provides a volatile market with high jumps reported (Catania and Sandholdt, 2019) that trades continuously: its daily boundary is a calendar convention rather than an exchange close. This thesis asks whether intraday jump timing improves one-day forecasts of its realized variance and whether a richer Hawkes process adds predictive value. The literature below establishes the objects and comparisons needed to make those questions precise.

1.2 Realized Volatility Forecasting

Realized variance (RV), defined as the sum of squared intraday returns, estimates daily quadratic variation under standard assumptions (Andersen et al., 2001b; Andersen et al., 2003). The resulting series exhibits several stylized facts commonly associated with volatility measures (Cont, 2001). Volatility clusters: high- and low-volatility periods tend to be followed by periods of similar volatility. RV also displays slowly decaying dependence, which has motivated the use of long-memory models in volatility forecasting. Its distribution is right-skewed and becomes more nearly normal after a logarithmic transformation. Volatility also responds asymmetrically to return signs: negative returns tend to raise volatility more than positive returns do (Andersen et al., 2001a; Cont, 2001). For this research, the central feature is that quadratic variation contains both continuous and discontinuous variation due to jumps. This motivates methods that model these components separately (Andersen et al., 2007a). RV estimation

also presents a sampling-frequency trade-off: finer sampling improves convergence to quadratic variation under ideal conditions, but makes the estimator more susceptible to market microstructure noise (Peter R. Hansen and Lunde, 2006; Liu et al., 2015). Forecast evaluation should therefore account for the properties of the realized measure, including through loss functions such as mean squared error (MSE) and quasi-likelihood (QLIKE) (Patton, 2011).

The HAR model of Corsi (2009) forecasts realized variance from daily, weekly, and monthly averages, capturing slow dependence with a short linear regression. Its performance makes it a demanding benchmark for extensions (Chassot and Audrino, 2026). Earlier ARCH and GARCH models instead infer conditional variance from daily return innovations (Engle, 1982; Bollerslev, 1986). Mixed-data sampling models use weighted intraday information to forecast at a lower frequency (Ghysels et al., 2006; Andreou, 2016); they provide a second route beside HAR rather than a choice made here.

HAR applications depend on several choices, as is common in forecasting: the forecast horizon, the calendar used to define weekly and monthly averages, the estimation window, the information set, the transformation of realized variance, the estimation method, and the loss function used for evaluation (Clements and Preve, 2021; Patton and Zhang, 2026). A forecaster must therefore specify these choices to define the forecasting problem and interpret its results.

HAR has been extended in several directions beyond the continuous–jump decomposition. Bipower variation helps identify jump variation and motivates separating continuous and jump components (Andersen et al., 2007a). Semivariances distinguish positive from negative return variation, allowing researchers to examine how "good" and "bad" volatility affect future volatility (Barndorff-Nielsen et al., 2008; Corsi and Renò, 2012). Measurement-error corrections address noise in realized-variance regressors (Y. Wang et al., 2020). Flexible predictors have also improved on HAR in some equity samples (Christensen et al., 2023), although evidence on machine-learning extensions is mixed and is beyond the scope of this discussion.

A central limitation of using the jump component to forecast variance is that daily aggregation does not reveal when jumps occurred or whether individual intraday returns are statistically significant jumps. For daily jump detection, Huang and Tauchen (2005) provide a useful overview of statistics for the Barndorff-Nielsen–Shephard (BNS) test (Barndorff-Nielsen and Shephard, 2006), which can identify jump days at a chosen significance level. Intraday methods, by contrast, typically identify unusually large returns, with procedures differing in how they

standardize returns and account for intraday variation (Andersen et al., 2007b; S. S. Lee and Mykland, 2008; Andersen et al., 2010; Clements and Liao, 2017).

1.3 Cryptocurrency Market and Bitcoin Volatility

Cryptocurrency markets trade throughout the week and across venues, making volatility relevant to positions whose risk does not pause at an exchange close. A daily boundary must therefore be chosen, and an observed price series represents a particular venue and trading pair rather than a consolidated market. Overlapping regional activity and funding times create within-day variation in trading conditions (Peter Reinhard Hansen et al., 2024; J.-N. Wang et al., 2020).

Bitcoin was the first cryptocurrency (Nakamoto, 2008) and occupies a central position in the market. Studies document return and volatility spillovers between Bitcoin and other crypto assets, including Ethereum and Litecoin (Koutmos, 2018; Kyriazis, 2019). Their relatedness matters when reading evidence across assets: similar findings do not constitute independent replications.

Bitcoin's realized volatility closely shares the stylized facts stated above for realized volatility: it shows persistence, intraday periodicity, asymmetry, and discontinuous movements (Catania and Sandholdt, 2019; Baur and Dimpfl, 2018). HAR forecasts it well relative to daily-return volatility models, particularly at short horizons (Aalborg et al., 2019; Bergsli et al., 2022). Studies extend its cascade with measurement-error adjustments, seven- and thirty-day windows, volume, tail measures, durations, and structural breaks (Qiu et al., 2021; Shen et al., 2020; Gkillas et al., 2021; Slim et al., 2023).

Large Bitcoin returns are frequent enough that jump detection and the information carried by jumps require particular care. Catania and Sandholdt (2019) find a high share of jump days but little effect of lagged jumps on Bitcoin volatility after the usual predictors are included. The distinction between jump size and jump arrival is therefore consequential: a weak daily jump-variation coefficient does not answer whether the timing and clustering of detected intraday events forecast variance.

1.4 Hawkes Process Applications

A Hawkes process models events whose occurrence raises the rate of subsequent events (Hawkes, 1971) and therefore presents clustering. Its branching interpretation separates background arrivals from events triggered by earlier ones (Hawkes and Oakes, 1974). The same structure has been used in several domains, such as earthquake aftershocks, crime, epidemics, and information diffusion (Yoshiko Ogata, 1988; Mohler et al., 2011; Reinhart, 2018; Rizoiu et al., 2017). Financial applications are also explored and detailed in Emmanuel Bacry et al. (2015). Among these applications, though not limited to them, are modeling financial trade arrivals for market microstructure studies and detecting and modeling extreme returns to produce superior tail-risk measure forecasts.

Several other applications to finance can be mentioned, such as studies of market reflexivity, clusters of extreme returns, and credit events (Filimonov and Sornette, 2012; Hardiman et al., 2013; Chavez-Demoulin et al., 2005; Errais et al., 2010). Nevertheless, all of them have the same challenges in common: their fitted excitation and memory depend on the event definition, estimation window, and kernel selection. An exponential kernel is computationally convenient, but an assumed decay shape can materially change fitted branching ratios (Filimonov and Sornette, 2015; Hardiman et al., 2013) and may fail to capture long-term memory dynamics (Emmanuel Bacry et al., 2015).

Under an exponential kernel, several modeling choices remain. In the basic unmarked model, every event contributes the same excitation, which then decays at a single rate. Marked extensions allow event characteristics to affect the strength of subsequent excitation: for example, an event's size can influence its impact on the intensity (Fuentes et al., 2018; Qu et al., 2024). Sum-of-exponentials kernels combine several decay rates, allowing excitation to operate over multiple time scales (Lallouache and Challet, 2016). These extensions capture distinct features of event clustering; whether they provide a better description depends on the event definition, data, and estimation window.

1.5 Limits in the Literature

Three of the closest papers are used to map the limits of the investigation. Clements and Liao (2017) introduce the use of intraday and daily equity jump and cojump intensities in a HAR forecasting setting, stating that these signals improve in- and out-of-sample performance, particularly in the days following jumps. Qu

et al. (2024) propose a step forward in the daily setting by incorporating a marked Hawkes process with a particular mark design to modulate the jumps' influence by different aspects of volatility. They show the potential of richer Hawkes models to yield out-of-sample statistical and economic gains. Pierre (2024), in a doctoral thesis, explores this intraday signal, fits two-decay Hawkes processes to intraday jumps in perpetual futures, and reports forecast gains against HAR.

These studies establish the potential of setting Hawkes jump intensities in HAR regressions, but some comparisons and questions remain unresolved, to the best of my knowledge. For example, Qu et al. (2024) limit their study to the daily setting without considering intraday jumps. It is also unclear whether a better fit produces better forecasts and how to fully specify a regression with this type of intensity, even though Pierre (2024) explores estimation methods. Additionally, although Clements and Liao (2017) compare the Hawkes signal with jump count measures, the comparison is not fair because the Hawkes signal takes into consideration a long window of jumps and uses fitted parameters for their decay structure, while the count benchmark does not. This leaves unresolved the question of whether the Hawkes machinery is necessary to build the intensity or whether simpler versions can perform equally well or better. Finally, the consensus indicates that the gains of these models come from modeling the jumps, but it has not been studied whether a volatility measure based on the same history and with similar properties can perform equally well or better without using jumps, or whether the jump gains persist after such control measures.

This investigation aims to address those questions by using the BTC market as an empirical example because of its accessibility to intraday data through Binance and the market properties described previously.

1.6 Investigation Statements

The general objective specifies the forecasting question. The specific objectives separate construction of the signal, its matched evaluation, the relation between event fit and forecast performance, and the necessity of Hawkes machinery. The contributions then state what the completed comparisons establish within their tested scope.

General objective

Evaluate whether information about the timing and clustering of intraday jumps improves Bitcoin realized-volatility forecasts.

Specific objectives

Five objectives organize the investigation:

1. Design a clear forecasting framework for applying Hawkes processes to intraday jumps in HAR forecasting.
2. Find a coherent regressor to represent the jump-intensity signal in the regression.
3. Study whether improvements in the fit of richer exponential Hawkes processes also improve out-of-sample volatility forecasts.
4. Test whether Hawkes estimation is needed for the forecast gains from jumps by comparing its regressor with a no-Hawkes decayed event count tuned on the same jumps.
5. Evaluate whether intraday jump intensities survive when a realized measure built on all returns with a similar decay structure is incorporated as an additional regressor.

1.7 Main Contributions and Thesis Outline

The thesis finds that the timing of intraday jumps improves one-day Bitcoin realized-volatility forecasts relative to HAR-CJ under log MSE and HRMSE, but that recent volatility measured from all returns carries most of that information. Its contributions follow the specific objectives:

1. A matched forecasting framework for intraday Hawkes regressors, in which parameters and regressor histories are recomputed at every origin and all models share forecast origins and information sets (Chapter 2 and section 3.2).
2. The conditional expected count of the next day as the jump-intensity regressor. It is affine in the closing intensity for the exponential process and is extended to the marked two-exponential process; the closing state

improves the benchmark where daily and integrated intensities do not (Chapter 2, appendix A, and section 3.2).

3. Evidence that a better fit to jump arrivals does not imply a better forecast: a second exponential is preferred by the likelihood at every refit without a significant forecast gain (Section 3.3).

4. Evidence that Hawkes estimation is not necessary for the gain: a decayed jump count with forecast-tuned memory is equally competitive (Section 3.2).

5. Evidence that the intensity adds no significant MSE improvement beside recent all- return variance, whose own gain is about ten times larger (Section 3.4).

These results hold for the tested event sets, one-day horizon, losses and benchmarks. Ethereum and Litecoin share only the HRMSE gains, no jump regressor improves seven- or thirty-day forecasts, and the economic exercises show no advantage. An insignificant difference is not equivalence, and model-confidence-set retention depends on the candidate pool and configuration.

Chapter 2 defines the realized measures, jump tests, event processes, regressors and forecast evaluation. Chapter 3 presents the Bitcoin data and forecast comparisons, the all-return controls, the robustness checks and the Ethereum and Litecoin results. Chapter 4 answers the research question and states its limits. The appendices contain the derivations, the simulation validation, the S&P 500 application, supplementary tables and reproducibility details.

Chapter 2

Theory and Methods

This chapter states the methods that Chapter 3 applies, in the order in which they are used. Realized measures and jump tests define the forecasting target and the events (Section 2.1). A point process models the arrival of those events (Section 2.2). It is estimated by maximum likelihood and simulated (Section 2.2.4), and compared with richer processes (Section 2.2.6). The fitted process is reduced to a daily regressor (Section 2.3), which enters a forecasting regression (Section 2.4), and the forecasts are evaluated (Section 2.5).

2.1 Realized Volatility and Jump Measures

Realized measures define the variance to forecast and separate its continuous and jump components. Statistical jump tests then select the event times to which the point processes are fitted. The amount of jump variation and the occurrence of a detected event are distinct inputs.

2.1.1 Quadratic variation and realized measures

A semimartingale log price accommodates continuous trading gains and discontinuous price changes in the usual no-arbitrage setting (Delbaen and Schachermayer, 1994; Andersen et al., 2003). Write $p(t) = 100 \log P(t)$, so price changes are in percentage units. The working finite-activity model is

$$\mathrm{d}p(t) = a(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t) + \kappa(t)\,\mathrm{d}J(t), \tag{2.1}$$

where W is Brownian motion, J counts jumps, a is locally bounded, and the positive volatility process satisfies the regularity conditions for realized variation

(Barndorff-Nielsen and Shephard, 2004).

Day t denotes $(t-1, t]$ on the market's daily clock. Its quadratic variation is the sum of integrated continuous variance and squared jump sizes:

$$QV_t = \underbrace{\int_{t-1}^{t} \sigma^2(s)\,\mathrm{d}s}_{IV_t} + \sum_{t-1<s\le t} \kappa^2(s). \tag{2.2}$$

Both terms use squared percentage units because p is already scaled. This decomposition defines the population quantities estimated below (Barndorff-Nielsen and Shephard, 2004).

For M equally spaced observations within day t, let $r_{t,i} = p_{t,i} - p_{t,i-1}$. Realized variance estimates quadratic variation:

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2 \xrightarrow{p} QV_t \qquad (M \to \infty). \tag{2.3}$$

The limit equals integrated variance in the absence of jumps; the convergence assumes that observational noise does not dominate the sampling limit (Andersen et al., 2003).

Bipower variation estimates integrated variance even with finite-activity jumps (Barndorff-Nielsen and Shephard, 2004). The staggered version used by Clements and Liao (2017) is

$$BPV_t = \frac{\pi}{2} \frac{M}{M-2} \sum_{i=3}^{M} |r_{t,i}|\,|r_{t,i-2}| \xrightarrow{p} IV_t. \tag{2.4}$$

Skipping the adjacent return reduces exposure to noise shared by consecutive returns; the consistency statement still requires the cited sampling conditions.

The forecasting decomposition uses bipower variation for the continuous component and truncates the estimated jump contribution at zero:

$$C_t = BPV_t, \qquad J_t = \max\{RV_t - BPV_t, 0\}. \tag{2.5}$$

Thus C_t estimates integrated variance and J_t estimates squared jump variation. This statistical decomposition need not add exactly to RV_t in finite samples when $BPV_t > RV_t$; it also differs from selecting a set of statistically significant jump events.

2.1.2 Jump detection

A positive $RV_t - BPV_t$ can arise from sampling variation without a jump. BNS tests standardize that difference using its null variance; difference, ratio, and transformed statistics offer related ways to test the no-jump hypothesis (Barndorff-Nielsen and Shephard, 2006; Huang and Tauchen, 2005). The ratio form below measures the discrepancy relative to the day's realized variance.

Daily detection asks whether at least one jump occurred during the day; intraday detection locates the returns carrying discontinuities. The daily BNS rule defines jump-day events, whereas ABD, Lee–Mykland, and sequential BNS identify intraday event times. Their combinations provide alternative event sets for the same forecasting question.

The daily test: BNS with the ratio statistic

The uncertainty of the RV–BPV difference depends on integrated quarticity. A jump-robust estimate uses tripower products (Barndorff-Nielsen and Shephard, 2004):

$$TQ_t = Mc_{4/3}^{-3}\frac{M}{M-4}\sum_{i=5}^{M}|r_{t,i-4}|^{4/3}|r_{t,i-2}|^{4/3}|r_{t,i}|^{4/3}, \tag{2.6}$$

where $c_{4/3} = \mathbb{E}|Z|^{4/3}$ for $Z \sim N(0,1)$. Under the finite-activity assumptions, it estimates $\int_{t-1}^{t}\sigma^4(s)\,\mathrm{d}s$ in fourth powers of percentage units.

With $\Delta = 1/M$, the ratio statistic and jump-day decision are

$$Z_t^{BNS} = \frac{(RV_t - BPV_t)/RV_t}{\sqrt{\Delta(\pi^2/4+\pi-5)\max\{1, TQ_t/BPV_t^2\}}}, \qquad Z_t^{BNS} > \Phi^{-1}(1-\alpha). \tag{2.7}$$

The null limit is standard normal (Barndorff-Nielsen and Shephard, 2006). Scaling the discrepancy by RV_t makes it relative to the day's variation; rejection identifies a jump day but does not locate its jumps.

The intraday test of Andersen, Bollerslev and Dobrev

The ABD rule flags an intraday return when its absolute size exceeds a daily volatility-scaled threshold (Andersen et al., 2007b):

$$|r_{t,i}| > \Phi^{-1}(1-\pi_M/2)\sqrt{BPV_t/M}, \qquad \pi_M = 1-(1-\alpha_{ABD})^{1/M}. \tag{2.8}$$

Write $\mathcal{J}_t^{ABD}$ for the flagged indices.

The scale $\sqrt{BPV_t/M}$ represents a typical continuous return within the day, so volatile days require larger absolute returns to trigger detection. The Šidák adjustment translates the chosen daily significance level into a per-return threshold under the test's null approximation. It does not guarantee exact daily size when the standardized returns remain dependent.

A local normalization: the Lee–Mykland test

The Lee–Mykland test uses a local volatility estimate (S. S. Lee and Mykland, 2008). With returns indexed consecutively by j, define

$$L_{t,i} = r_{t,i}/\hat{\sigma}_{t,i}, \qquad \hat{\sigma}^2_{t,i} = \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} |r_j|\,|r_{j-1}|. \tag{2.9}$$

For $c = \sqrt{2/\pi}$ and $g = -\log[-\log(1-\alpha_{LM})]$, its flagged set $\mathcal{J}_t^{LM}$ satisfies

$$|L_{t,i}| > \frac{g}{c\sqrt{2\log M}} + \frac{\sqrt{2\log M}}{c} - \frac{\log\pi + \log\log M}{2c\sqrt{2\log M}}. \tag{2.10}$$

The local bipower estimate adapts to recent volatility rather than holding one scale throughout the day. The threshold follows the limiting extreme-value law of the maximum standardized return and uses a daily significance level. The lookback K controls how quickly the scale changes; its empirical value and intraday normalization are recorded in Appendix E.

The sequential BNS test

Sequential BNS locates intraday jumps by repeatedly applying the daily statistic (Andersen et al., 2010; Clements and Liao, 2017). Compute Z_t^{BNS}; if it rejects, flag the largest standardized absolute return, replace that return by zero, and recompute RV_t, BPV_t, and TQ_t. Repeat until the statistic no longer rejects.

Removing the largest flagged return allows the test to find further discontinuities that were masked by the first one. A day on which the initial daily test does not reject supplies no sequential events. The resulting set $\mathcal{J}_t^{sBNS}$ therefore shares its event days with the daily test but can contain several returns within a day.

Combining intraday jump sets

The three intraday tests differ in their volatility scale and detection mechanism: ABD uses a daily scale, Lee–Mykland a local scale, and sequential BNS repeated daily tests. They need not flag the same returns. Their disagreement motivates combining event sets, with different tradeoffs between missed jumps and spurious detections (Dumitru and Urga, 2012).

The intersection retains flags shared by all three tests, while the union retains any flag:

$$\mathcal{J}_t = \mathcal{J}_t^{ABD} \cap \mathcal{J}_t^{LM} \cap \mathcal{J}_t^{sBNS}, \qquad \mathcal{U}_t = \mathcal{J}_t^{ABD} \cup \mathcal{J}_t^{LM} \cup \mathcal{J}_t^{sBNS}. \tag{2.11}$$

Event times are the closing instants of the retained returns. Agreement produces a conservative event definition without ensuring that every flag is a true jump; the union increases coverage and may admit more false flags. Chapter 3 fixes the primary set and evaluates alternatives.

2.2 Temporal Point Processes

Jump detection produces event times. A temporal point process models their conditional arrival rate given the observed history. The Poisson model provides a reference without event-driven excitation; Hawkes models let past events alter that rate.

2.2.1 Conditional intensity and Poisson Process

Let $N(A)$ count events in interval A and let $\mathcal{H}_{s^-}$ contain events and marks strictly before s. The predictable conditional intensity is the expected rate of new arrivals given that history (Daley and Vere-Jones, 2003):

$$\lambda^*(s) = \lim_{\epsilon \downarrow 0} \frac{\mathbb{E}\left[N([s, s+\epsilon)) \mid \mathcal{H}_{s^-}\right]}{\epsilon}. \tag{2.12}$$

For small $\epsilon > 0$, the expected count on $[s, s+\epsilon)$ is therefore $\lambda^*(s)\epsilon + o(\epsilon)$. Its integral is the compensator

$$\Lambda^*((a, b]) = \int_a^b \lambda^*(s)\, \mathrm{d}s. \tag{2.13}$$

The intensity assigns rates to observed events in the likelihood, while its integral accounts for exposure over the whole observation window. The compensated process $N - \Lambda^*$ is a martingale under the integrability conditions used here (Daley and Vere-Jones, 2003). This connects expected counts to integrated predictive intensity and supplies the time-rescaling diagnostics in Appendix B.4.

A Poisson process has independent counts on disjoint intervals and a deterministic intensity $\lambda^*(s) = \lambda(s)$ (Daley and Vere-Jones, 2003). It is homogeneous when that rate is constant and inhomogeneous when it varies with time. Its expected count on $(a, b]$ is $\int_a^b \lambda(s)\,\mathrm{d}s$, with no dependence on the realized event history.

An inhomogeneous Poisson rate can describe predictable calendar variation, but cannot make an observed jump raise the subsequent arrival rate. That limitation matters when the event history contains additional clustering (Clements and Liao, 2017; Qu et al., 2024). Testing Hawkes against Poisson asks whether event-dependent excitation improves the description; it does not identify a causal source of the observed clustering.

2.2.2 The Hawkes process

The linear Hawkes process adds nonnegative excitation from earlier events to a baseline (Hawkes, 1971):

$$\lambda^*(t) = \mu(t) + \sum_{t_q<t} \phi(t - t_q), \qquad \phi \geq 0. \tag{2.14}$$

An event raises the rate through ϕ, and $\phi \equiv 0$ recovers a Poisson process with rate $\mu(t)$, inhomogeneous unless μ is constant.

In the branching representation, baseline arrivals are immigrants and each event generates offspring at ages governed by ϕ (Hawkes and Oakes, 1974). The integrated kernel $\eta = \int_0^\infty \phi(u)\,\mathrm{d}u$ is the expected number of direct offspring, while its decay describes memory. This is a probabilistic decomposition of the fitted process; it does not identify external and endogenous causes of individual financial jumps.

For constant $\mu > 0$, the unmarked exponential model is

$$\lambda^*(t) = \mu + \alpha\beta \sum_{t_q<t} e^{-\beta(t-t_q)}, \qquad \alpha \geq 0,\ \beta > 0. \tag{2.15}$$

Its excitation state decays at rate β between events and rises by $\alpha\beta$ at an event.

This single state is Markovian and can be updated recursively, giving linear-time likelihood evaluation. Its post-event value at the daily close contains the history needed for the conditional count derived below.

2.2.3 Richer exponential self-exciting

A marked point process attaches a positive characteristic ξ_q to event t_q (Daley and Vere-Jones, 2003). The power-marked bi-exponential specification is

$$\lambda^*(t) = \mu + \sum_{t_q<t} \xi_q^{\delta} \sum_{k=1}^{2} \alpha_k \beta_k e^{-\beta_k (t-t_q)}, \qquad \beta_1 > \beta_2 > 0, \quad \alpha_k, \delta \geq 0. \tag{2.16}$$

Power marks allow event characteristics to change excitation; the two decay rates allow fast and slow memories.

Setting $\alpha_2 = 0$ recovers a marked exponential model; setting $\delta = 0$ removes the marks; imposing both gives (2.15). Setting both amplitudes to zero gives homogeneous Poisson. The one-component notation used below is

$$\lambda^*(t) = \mu + \sum_{t_q<t} \alpha\beta\xi_q^{\delta} e^{-\beta(t-t_q)}. \tag{2.17}$$

Here $\alpha\xi^{\delta}$ is the integrated event impact and $\beta e^{-\beta u}$ its normalized decay.

The daily MI model of Qu et al. (2024) uses median realized variance as a jump-day mark. For intraday returns,

$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \frac{M}{M-2} \sum_{i=2}^{M-1} \text{med}(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}|)^2 \tag{2.18}$$

is the jump-robust measure of Andersen et al. (2012). Its persistent financial mark requires care when interpreting branching quantities based on independent marks. If the conditional mean of the mark impact at an arrival is a constant $\nu_\delta = \mathbb{E}[\xi^\delta]$, the offspring mean is

$$\eta = (\alpha_1 + \alpha_2)\nu_\delta, \qquad \lambda_\infty = \mu/(1-\eta), \quad \eta < 1. \tag{2.19}$$

Independent identically distributed marks independent of preceding events are sufficient. Under the usual existence conditions, $\eta < 1$ gives a stationary process with finite mean intensity (Daley and Vere-Jones, 2003). One component sets $\alpha_2 = 0$; unmarked events have $\nu_0 = 1$. A sample mark moment alone does not establish the conditional-moment assumption.

The time h_k at which one exponential contribution halves satisfies $e^{-\beta_k h_k} = 1/2$, hence $h_k = \log 2/\beta_k$. The richer kernel therefore has a fast and a slow component half-life, not a single half-life of their sum. Each component has its own Markov excitation state; retaining both is necessary for its expected-count calculation and efficient likelihood evaluation.

2.2.4 Maximum likelihood estimation

Maximum likelihood fits the conditional rate of the observed events. Following Daley and Vere-Jones (2003), the star in λ^* indicates conditioning on the strict past. For marked events, $\lambda^*(t, \xi)$ specifies a rate jointly over time and mark, while $\lambda^*(t)$ denotes the ground intensity obtained by integrating over marks.

The likelihood and its evaluation

For events $\{(t_q, \xi_q)\}_{q=1}^n$ on $[0, T]$ with marks in $\mathcal{X} = (0, \infty)$, the likelihood combines the rates at observed events with the compensator over the entire window and over all marks (Daley and Vere-Jones, 2003):

$$\begin{aligned} L &= \prod_{q=1}^{n} \lambda^*(t_q, \xi_q) \exp\left[-\int_0^T \int_{\mathcal{X}} \lambda^*(s, \xi) \, \mathrm{d}\xi \, \mathrm{d}s \right] \\ &= \prod_{q=1}^{n} \lambda^*(t_q, \xi_q) \exp\left[-\int_0^T \lambda^*(s) \, \mathrm{d}s \right]. \end{aligned} \tag{2.20}$$

The inner integral is the ground intensity, $\int_{\mathcal{X}} \lambda^*(s, \xi) \, \mathrm{d}\xi = \lambda^*(s)$, so the exponent integrates over marks as well as time. The event-free interval after the final event remains part of the observation window and must enter the integral.

Factor the marked intensity into the ground intensity and the conditional density of the mark at an arrival,

$$\lambda^*(t, \xi) = \lambda^*(t; \theta) \, f(\xi \mid \mathcal{H}_{t^-}; \psi), \qquad \int_{\mathcal{X}} f(\xi \mid \mathcal{H}_{t^-}; \psi) \, \mathrm{d}\xi = 1. \tag{2.21}$$

The mark enters each event's contribution through f. Because f integrates to one over $\mathcal{X}$, the exponent of (2.20) depends on θ only, and the log-likelihood is

$$\log L = \underbrace{\sum_{q=1}^{n} \log \lambda^*(t_q; \theta) - \int_0^T \lambda^*(s; \theta) \, \mathrm{d}s}_{-\mathrm{NLL}_g(\theta)} + \underbrace{\sum_{q=1}^{n} \log f(\xi_q \mid \mathcal{H}_{t_q^-}; \psi)}_{\text{mark part}}. \tag{2.22}$$

When θ and ψ vary freely and share no parameter, the mark-density likelihood

separates, so fitting θ minimizes the ground negative log-likelihood, with subscript g for ground,

$$\mathrm{NLL}_g(\theta) = \int_0^T \lambda^*(s;\theta)\,\mathrm{d}s - \sum_{q=1}^{n} \log \lambda^*(t_q;\theta). \tag{2.23}$$

This separability permits fitting the event-time parameters without estimating the mark density; it does not make the marks independent.

The estimate solves $\hat{\theta} = \arg\min_{\theta\in\Theta} \mathrm{NLL}_g(\theta)$ with positive baseline and decays, nonnegative amplitudes and mark exponent, and ordered decays for two components. The objective is nonconvex, so the numerical search uses several starting decay rates and retains the best converged optimum. The empirical solver bounds the longest component half-life by the observation window and does not impose stationarity. Appendix E records the starts, constraints, convergence rules, and handling of boundary solutions.

For the marked two-state model, let $x_k(s) = \alpha_k\beta_k \sum_{t_q<s} \xi_q^\delta e^{-\beta_k(s-t_q)}$ be the excitation of component k from events strictly before s, and $S_k(q) = x_k(t_q)$ its value at the qth event. Event times are taken to be distinct, so the strict past of t_q contains exactly the events $1,\dots,q-1$. With an empty pre-window history, $S_k(1) = 0$ and

$$S_k(q) = e^{-\beta_k(t_q - t_{q-1})}\left[S_k(q-1) + \alpha_k\beta_k\xi_{q-1}^\delta\right], \qquad k = 1, 2. \tag{2.24}$$

Then $\lambda^*(t_q) = \mu + S_1(q) + S_2(q)$. Only the preceding event enters the update, preserving predictability. This exponential recursion evaluates both states in one pass (Yosihiko Ogata, 1981); setting one amplitude to zero gives the one-state algorithm.

Algorithm 1 Ground log-likelihood of the power-marked two-exponential process

Require: ordered marked events on $[0, T]$ and parameters of (2.16)
1: $S_1, S_2 \leftarrow 0$; $\ell \leftarrow 0$; $\Lambda \leftarrow \mu T$
2: **for** $q = 1,\dots,n$ **do**
3: **if** $q > 1$ **then**
4: update both states by (2.24)
5: **end if**
6: $\ell \leftarrow \ell + \log(\mu + S_1 + S_2)$
7: $\Lambda \leftarrow \Lambda + \sum_{k=1}^{2} \alpha_k\xi_q^\delta(1 - e^{-\beta_k(T-t_q)})$
8: **end for**
9: **return** $\ell - \Lambda$

Each component also has an analytic integrated kernel, so the same pass

evaluates

$$\mathrm{NLL}_g = \mu T + \sum_{q=1}^{n} \sum_{k=1}^{2} \alpha_k \xi_q^{\delta} (1 - e^{-\beta_k (T - t_q)}) - \sum_{q=1}^{n} \log(\mu + S_1(q) + S_2(q)). \quad (2.25)$$

There are two state updates and two integrated contributions per event, giving $O(n)$ cost. The empty-history convention can create an edge effect, assessed in Appendix B.5; the analytic gradient of the unmarked exponential case is derived in Appendix A.1.

Inference on the estimates

For an interior identifiable parameter in a stationary process satisfying the regularity conditions of Yoshiko Ogata (1978), $\sqrt{T}(\hat{\theta} - \theta^{\circ}) \Rightarrow N(0, \mathcal{I}(\theta^{\circ})^{-1})$, where $\mathcal{I}$ is information per unit time. The inverse observed Hessian of the full-window negative log-likelihood estimates the covariance of $\hat{\theta}$. Boundary optima, weak identification, and finite event counts can invalidate this approximation; such standard errors are not interpreted as regular interior inference. Simulation checks are reported in Appendix B.5.

2.2.5 Simulation

Point-process paths can be generated by inversion of integrated intensity, by branching constructions, or by thinning a dominating process (Yosihiko Ogata, 1981; Hawkes and Oakes, 1974). The choice depends on whether the conditional waiting-time law, offspring distribution, or a usable intensity bound is easiest to evaluate.

Ogata's modified thinning draws a candidate from a rate bounding the current intensity, accepts it with probability equal to the intensity-to-bound ratio, and updates the bound after each candidate (Yosihiko Ogata, 1981). Exponential Hawkes states decay between events, making their current post-event intensity a bound until the next candidate. Accepted events update both states and, for a marked model, carry a mark drawn from the specified generating law. Appendix B gives the algorithm.

Simulation supplies controlled comparisons against known parameters and conditional means. It checks whether the likelihood is evaluated correctly, whether estimators recover generating values, and whether nominal tests reject at the intended rate. It also approximates finite-sample reference distributions when standard likelihood-ratio or goodness-of-fit limits do not apply.

In this investigation simulation principally supplies bootstrap reference distributions. Additional exercises calibrate the estimator, validate the simulator, and verify expected counts. Appendix B identifies the generating process and evidence available for each check; success on simulated paths does not establish that observed financial jumps follow that process.

2.2.6 Likelihood ratio test

Goodness-of-fit asks whether a fitted model adequately describes the observed event times; model comparison asks whether one specification describes them better than another. Time-rescaling diagnostics and their refitted bootstrap are developed in Appendices B.4 and B.6. A relative likelihood advantage can coexist with rejection of both models by a goodness-of-fit test.

Information criteria penalize fit by complexity, predictive scores assess held-out events, and likelihood ratios compare nested restrictions. The nested model ladder here uses the latter: fit both models on the same events and window and compute

$$LR = 2(\ell_1 - \ell_0), \tag{2.26}$$

where ℓ_1 and ℓ_0 are the maximized log-likelihoods of the larger model and of its nested restriction. At exact maxima this is nonnegative; a negative numerical ratio is reported as an optimization failure rather than clipped. Daily and intraday event likelihoods are not compared directly.

For k interior, identified restrictions under regular likelihood conditions, $LR \Rightarrow \chi^2_k$. A single boundary restriction can give the mixture $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ under the corresponding regularity assumptions (Self and Liang, 1987). That reference is used for a nonnegative mark exponent at zero. If an added decay is unidentified under the null, as for Poisson versus Hawkes or one versus two exponentials, these simple laws need not apply (Davies, 1987).

For the nonstandard comparisons, simulate from the fitted null, refit both models on each path, and use the empirical upper quantile of the resulting likelihood ratios. The reference must reproduce the observation grid and fitting procedure being assessed. Appendix B.9 gives the bootstrap procedure and distinguishes available calibration evidence from comparisons whose finite-sample size has not been established.

2.3 From a Fitted Process to a Daily Regressor

A continuous-time event model must supply a daily input to the forecasting regression. At the close of day t, let $\lambda_t = \lambda^*(t^+)$ include any event on the final observed return. The target summary is $\mathbb{E}\Lambda_t(h) = \mathbb{E}[N((t, t+h]) \mid \mathcal{H}_t]$, the expected event count over the next h days.

2.3.1 The expected count over a forecast window

The conditional expected count is the integral of the mean predictive intensity. For the power-marked two-state model, put $x_t = (x_1(t^+), x_2(t^+))^\top$, $a = (\alpha_1\beta_1\nu_\delta, \alpha_2\beta_2\nu_\delta)^\top$, and $A = -\operatorname{diag}(\beta_1, \beta_2) + a\mathbf{1}^\top$. Then

$$\mathbb{E}\Lambda_t(h) = \int_0^h [\mu + \mathbf{1}^\top \bar{x}(u)]\,\mathrm{d}u, \qquad \bar{x}'(u) = A\bar{x}(u) + \mu a, \quad \bar{x}(0) = x_t. \tag{2.27}$$

This gives an affine function of the vector of closing states. In general their sum, the scalar intensity, is insufficient.

The result holds with fixed parameters over the forecast window, finite conditional first moments, and a constant conditional mean ν_δ of ξ^δ at an arrival. An independent mark distribution is sufficient; persistent volatility marks require a separate justification. The stationary representation additionally requires $\eta = (\alpha_1 + \alpha_2)\nu_\delta < 1$, whereas the finite-horizon state equation does not divide by $1 - \eta$. Events at the daily close are included in the conditioning state.

The derivation first applies the compensator martingale identity to future counts, then takes conditional expectations of each excitation-state equation. Solving the resulting linear system and integrating its total intensity yields the count. Appendix A.3 gives the full derivation, including a matrix-exponential form that remains well defined at the stationarity boundary.

For one unmarked exponential, $\eta = \alpha$ and $\beta_* = \beta(1 - \eta)$. When $\eta < 1$, the conditional mean intensity is $m_t(t + u) = \lambda_\infty + (\lambda_t - \lambda_\infty)e^{-\beta_* u}$, where $\lambda_\infty = \mu/(1 - \eta)$. Integration gives

$$\mathbb{E}\Lambda_t(h) = \lambda_\infty h + (\lambda_t - \lambda_\infty)\frac{1 - e^{-\beta_* h}}{\beta_*}. \tag{2.28}$$

The effective decay accounts for offspring generated within the forecast window. Appendix A.2 gives the derivation.

For fixed parameters and horizon, the expected count is affine in closing

intensity:

$$\mathbb{E}\Lambda_t(h) = a_h + b_h\lambda_t, \qquad b_h = \frac{1 - e^{-\beta_* h}}{\beta_*}, \quad a_h = \lambda_\infty(h - b_h). \tag{2.29}$$

With an intercept, replacing intensity by this count leaves a linear regression's fitted values and forecasts unchanged, provided the same transformation applies to every estimation and forecast row. Different parameter refits across rows or a logarithmic transformation need not preserve the equivalence.

Freezing history suppresses future offspring and gives $\mathbb{E}\Lambda_t^{(0)}(h) = \mu h + (\lambda_t - \mu)(1 - e^{-\beta h})/\beta$, the integrated deterministic decay used in the literature's daily signal (Clements and Liao, 2017). It is an inhomogeneous Poisson mean rather than the self-exciting conditional count. Both are affine in λ_t for fixed parameters, but their count interpretations differ.

The empirical regressors plug the latest parameter estimates into these expressions, using only information available at the forecast origin. Between refits, parameters remain fixed and new events update the states. Parameter uncertainty can be propagated to the count by the delta method; Appendix B.8 derives that approximation, including the dependence of the reconstructed state on the fitted parameters. It is presented as inference on the estimated count, not as a prediction interval for the realized future count or an uncertainty calculation already performed in the forecasting tables.

2.3.2 The regressors

The intensity summaries compare information available at the daily close. $EI_t = \lambda_t$ retains the intraday closing state; $LCI_t = \log \mathbb{E}\Lambda_t(1)$ uses its expected count on a log scale. The literature's $SI_t = \int_{t-1}^{t} \lambda^*(s)\,\mathrm{d}s$ aggregates the completed day's intraday intensity, while $DI_t = \lambda_D^*(t+1 \mid \mathcal{H}_t)$ is the next-day intensity from a process fitted to jump days (Clements and Liao, 2017; Qu et al., 2024). Richer processes use the corresponding fitted states, with their empirical transformations specified in the model taxonomy.

In levels, EI and the exact count are interchangeable under the common-parameter condition of Section 2.3.1; LCI is not an affine rescaling. The difference between EI and SI is temporal: an event near the close has little time to contribute to the day's integral but immediately raises the closing intensity. DI discards intraday ordering, so these regressors compare both aggregation and functional form.

A discounted event count uses the same timestamps without fitting a point process:

$$F_t(h_F) = \sum_{t_q \le t} 2^{-(t-t_q)/h_F}, \qquad \lambda_t = \mu + \alpha\beta F_t(\log 2/\beta). \tag{2.30}$$

The identity follows by substituting $h_F = \log 2/\beta$. Thus EI selects memory by event-time likelihood, whereas F selects its single half-life by the training regression's residual sum of squares. The daily counterpart DF uses jump days and a half-life of at least one day.

Closing intensity emphasizes recent events, motivating controls that measure late activity from every return:

$$EWRV_t(h_E) = \sum_{s \le t} r_s^2\, 2^{-(t-s)/h_E}, \qquad RV60_t = \sum_{s \in (t-1/24,\,t]} r_s^2. \tag{2.31}$$

Time is in days, h_E is tuned on the training regression, and both controls enter in logarithms. $EWRV$ tests the role of discounted memory; $RV60$ tests the final hour without a fitted decay. Incremental jump information is assessed against these stronger recent-volatility benchmarks.

2.4 Forecasting Realized Volatility

A HAR backbone forecasts log realized variance from daily, weekly, and monthly histories. Additional jump summaries test the information in event arrivals, and all-return controls test whether that information extends beyond recent variance. Every added regressor is observed by the forecast origin.

2.4.1 The HAR family

For a positive series X_t, write $\widetilde{X}_t = \log X_t$ and $\overline{\widetilde{X}}_t^{(k)} = k^{-1}\sum_{i=0}^{k-1} \widetilde{X}_{t-i}$. The HAR cascade averages logs, rather than taking the log of an arithmetic average:

$$\widetilde{RV}_t = \phi_0 + \phi_d \widetilde{RV}_{t-1} + \phi_w \overline{\widetilde{RV}}_{t-1}^{(w)} + \phi_m \overline{\widetilde{RV}}_{t-1}^{(m)} + \varepsilon_t. \tag{2.32}$$

Logarithms follow Clements and Liao (2017) and put the positive target on an unrestricted regression scale. Crypto calendars use $(1, 7, 30)$; equity literature configurations use $(1, 5, 22)$ sessions, with alternative cascades stated explicitly. The coefficients are fitted by ordinary least squares.

HAR-CJ is the primary benchmark, separating the continuous and jump

histories of (2.5) (Andersen et al., 2007a). Let $\mathcal{K} = \{1, w, m\}$, $\widetilde{C}_t = \log C_t$, and $\widetilde{J}_t = \log(1 + J_t)$. Then

$$\log RV_t = \phi_0 + \sum_{k\in\mathcal{K}} \phi_k \overline{\widetilde{C}}^{(k)}_{t-1} + \sum_{k\in\mathcal{K}} \gamma_k \overline{\widetilde{J}}^{(k)}_{t-1} + \varepsilon_t. \tag{2.33}$$

The shift by one keeps zero jump-variation days defined, while the bipower continuous component remains the positive measure specified in Chapter 3.

Signed and leverage variants supply alternative backbones: SHAR separates positive and negative variation, LHAR adds negative-return histories, and LSHAR-CJ combines signed and leverage terms with the jump cascade. Appendix A.4 gives their equations and relates the SHAR-CJ and LSHAR-CJ names to the recorded semivariance configurations. Each comparison holds the added regressor and forecast origins fixed.

2.4.2 HAR-H: a jump-arrival summary as the additional regressor

Write $B_{t-1}(\gamma)$ for the linear predictor, including its intercept, of a chosen HAR backbone. An arrival summary S_{t-1} gives

$$\log RV_t = B_{t-1}(\gamma) + \vartheta S_{t-1} + \varepsilon_t. \tag{2.34}$$

This extends the intensity-augmented regression of Clements and Liao (2017) to the summaries defined above. The backbone is recovered at $\vartheta = 0$.

A controlled comparison includes recent all-return variance A_{t-1}:

$$\log RV_t = B_{t-1}(\gamma) + \varrho \log A_{t-1} + \vartheta S_{t-1} + \varepsilon_t, \qquad A \in \{RV60, EWRV\}. \tag{2.35}$$

The null $\vartheta = 0$ now removes the jump summary from a backbone already containing recent volatility. This isolates its increment relative to the chosen control rather than its gain over daily aggregates alone.

Model names identify the backbone and added information: HAR-CJ-EI uses closing intraday intensity, HAR-CJ-LCI the log count, HAR-CJ-SI the completed-day integral, HAR-CJ-DI the daily-process intensity, and HAR-CJ-F the tuned discounted count. EI, SI, DI, and F enter in levels; LCI and the all-return controls enter in logs. DI/SI follow the literature, EI/LCI implement the conditional-count construction, and F tests the need for likelihood-based memory selection. The richer-process labels distinguish marked and two-component signals and are tied

to their exact transformations in Appendix E.

2.5 Forecast Evaluation

Forecast evaluation specifies the information available at each origin, a loss measuring its error, and inference for comparisons under dependence. Practical exercises then examine whether a statistical improvement changes an allocation or risk forecast. These stages answer distinct questions about the same stored forecasts.

2.5.1 Out-of-sample schemes

A forecasting scheme may use a fixed training sample, an expanding sample, or a rolling window of fixed length. Process parameters, forecasting coefficients, and tuned regressor half-lives may also be updated at different intervals. These choices trade adaptation against estimation noise and computational cost, and must be fixed before comparing models.

The selected scheme forecasts day t from information through day $t-1$ using a rolling window of W days. The regression is refitted at every origin; process parameters and tuned half-lives are updated at the specified cadence and held fixed between updates. New observed events still update the intensity state. Historical regressors are reconstructed using the active parameter vector so that the affine equivalence in Section 2.3.1 applies within each fit.

2.5.2 Loss functions

Realized variance is a noisy proxy for the latent variation of interest, so loss choice affects what a forecast ranking means (Patton, 2011). The comparison reports log-scale accuracy and level-scale volatility losses separately. Robustness properties proved for a loss on variance levels do not automatically transfer to squared errors of log variance.

Let $y_t = \log RV_t$ and $\hat{y}_t$ be its forecast. The level forecast is $\widehat{V}_t = \exp(\hat{y}_t + \hat{\sigma}_t^2/2)$, where $\hat{\sigma}_t^2$ is the training regression's residual variance. This lognormal correction adjusts for exponentiating a conditional mean in logs; its exact mean interpretation requires the associated distributional assumption.

The four losses are

$$L_{MSE,t} = (y_t - \hat{y}_t)^2, \qquad L_{MAE,t} = |y_t - \hat{y}_t|, \tag{2.36}$$

$$L_{QLIKE,t} = RV_t/\widehat{V}_t - \log(RV_t/\widehat{V}_t) - 1, \qquad L_{HRMSE,t} = (1 - \widehat{V}_t/RV_t)^2. \tag{2.37}$$

MSE emphasizes large log errors; MAE is less sensitive to them. QLIKE measures relative variance forecasting error and has the proxy-robust ranking property under the conditions of Patton (2011). HRMSE emphasizes proportional errors, particularly overprediction on quiet days; it is reported as the mean squared relative error without taking a square root.

2.5.3 Pairwise comparison

A lower average loss does not by itself establish a population accuracy difference. Daily loss differences are uncertain and may be serially dependent, especially with overlapping forecast targets. Tests must therefore address the sampling variation of the loss contrast rather than compare its sign alone.

Pairwise accuracy tests, nested-model corrections, and forecast-encompassing tests address complementary comparisons. The analysis uses Diebold–Mariano with the small-sample correction, Clark–West for an added regressor under squared loss, and encompassing tests for combinations of alternative forecasts (Diebold and Mariano, 1995; Harvey et al., 1997; Clark and West, 2007; Harvey et al., 1998).

For $d_t = L_t^{(1)} - L_t^{(2)}$, the Diebold–Mariano null is $\mathbb{E}[d_t] = 0$ and its statistic is

$$DM = \frac{\bar{d}}{\sqrt{\widehat{\Omega}_d/n}}, \qquad \widehat{\Omega}_d = \hat{\gamma}_0 + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\gamma}_j. \tag{2.38}$$

The implementation applies the small-sample adjustment of Harvey et al. (1997) and a two-sided Student reference with $n-1$ degrees of freedom. The horizon-specific lag convention and bootstrap sensitivity are stated separately.

For a nested benchmark forecast $\hat{y}_t^{(0)}$ and its extension $\hat{y}_t^{(1)}$, Clark–West adjusts squared errors for the larger model's estimation noise (Clark and West, 2007):

$$f_t = (y_t - \hat{y}_t^{(0)})^2 - (y_t - \hat{y}_t^{(1)})^2 + (\hat{y}_t^{(0)} - \hat{y}_t^{(1)})^2, \qquad CW = \sqrt{n}\bar{f}/\hat{\sigma}_f. \tag{2.39}$$

The comparison uses the upper normal tail and log squared error. Its adjusted

null differs from a direct comparison of achieved forecast losses.

Forecast encompassing asks whether the second forecast contributes to a combination with the first (Harvey et al., 1998). A descriptive regression is $y_t = a + w_1\hat{y}_t^{(1)} + w_2\hat{y}_t^{(2)} + u_t$. For errors e_1, e_2, the test uses

$$d_t = e_{1,t}(e_{1,t} - e_{2,t}), \qquad H_0 : \mathbb{E}[d_t] = 0, \quad H_1 : \mathbb{E}[d_t] > 0. \tag{2.40}$$

The DM construction is applied one-sided in both directions. Non-rejection in one direction combined with rejection in the reverse supports the reported encompassing interpretation; it does not prove the excluded forecast has no information.

The paired test measures realized loss differences, Clark–West addresses the nested extension after an estimation-noise adjustment, and encompassing asks about combining forecasts. Their conclusions need not coincide. Fixed-window comparisons and overlapping model specifications require the corresponding inferential qualifications (Giacomini and White, 2006); a significant nested test is not reported as a significant paired accuracy gain.

When a table contains a family of comparisons, Holm's step-down adjustment controls family-wise false rejection (Holm, 1979). Each table specifies its family, such as the four losses of one model pair. A benchmark comparison does not establish a difference between two extensions, and adjustment within one table does not control all hypotheses across the thesis.

2.5.4 Model confidence set

Several models can remain plausible when forecast differences are small. The model confidence set identifies models not eliminated by equal-predictive-ability tests within a stated candidate pool (Peter R. Hansen et al., 2011). It provides a set-valued comparison without forcing a unique winner.

Starting from the candidate set $\mathcal{M}$, test equal expected loss, remove the model selected by the statistic's elimination rule if the null is rejected, and repeat on the reduced set. Model p-values are cumulative maxima of elimination-step p-values; surviving models receive one. This preserves the ordering of evidence through elimination.

For $d_{ij,t} = L_{i,t} - L_{j,t}$ and $d_{i\cdot,t} = |\mathcal{M}|^{-1} \sum_j d_{ij,t}$, the two statistics and elimina-

tion rules are

$$T_R = \max_{i,j} |t_{ij}|, \qquad e_R = \arg\max_i \max_j t_{ij}, \tag{2.41}$$

$$T_{\max} = \max_i t_{i\cdot}, \qquad e_{\max} = \arg\max_i t_{i\cdot}, \tag{2.42}$$

where each t divides its mean contrast by its bootstrap standard error. Centered bootstrap contrasts supply the null distribution; the same index paths are reused and average contrasts are recomputed as the set shrinks.

Block resampling retains dependence within contiguous losses. MCS uses moving blocks; the paired-test sensitivity uses the stationary bootstrap with random block lengths (Politis and Romano, 1994). Both condition on stored forecasts and do not repeat model estimation or half-life tuning. Block-length sensitivity therefore assesses inference on these forecast sequences.

Report the candidate pool, loss, statistic, bootstrap settings, and retention threshold together. With B draws, the step p-value is $(1+\sum_b \mathbf{1}\{T_b^* \geq T\})/(B+1)$. Retain adjusted model p-values above the stated threshold. Joint retention is neither a proof of equality nor superiority; prevalence across overlapping configurations is descriptive rather than a collection of independent replications.

2.5.5 Horizons longer than one day

Portfolio and risk decisions may span several days, so an informative one-day regressor need not answer the practical longer-horizon question. Evaluating longer forecasts tests whether a short-lived arrival signal carries information beyond the next day.

For h days, the target is $y_t^{(h)} = \log(h^{-1} \sum_{i=0}^{h-1} RV_{t+i})$, forecast at the close of $t-1$. Exclude the final $h-1$ training rows whose targets are not yet observed. Consecutive targets overlap, inducing serial dependence; the reported paired statistic uses $q = h-1$ autocovariances, with dependence sensitivity interpreted under that forecast design.

2.5.6 Economic significance

Volatility forecasts influence exposure and risk limits, but a statistical loss reduction need not improve a decision. Economic value depends on the rule using the forecast, its costs, and the returns realized during evaluation.

Two illustrative exercises use the same forecasts: a capped risk-target allo-

cation and a standardized-return Value-at-Risk forecast. They assess practical consequences under specified rules and do not estimate an optimal trading policy. Detailed definitions and assumptions are collected in Appendix A.5.

With a daily variance forecast $\widehat{V}_t$ in decimal units, the risk-target weight is $w_t = \min\{1, \sigma^*/\sqrt{365\widehat{V}_t}\}$. Net portfolio returns are $R_t^p = w_t R_t - c|w_t - w_{t-1}|$, and annual certainty equivalent is $CE = 365(\bar{R}^p - \gamma s_R^2/2)$. These summarize the return-risk tradeoff of the rule; Appendix A explains its economic basis and trading simplifications.

If $r_t = \sqrt{V_t} Z_t$ with a stable standardized-return distribution, the conditional return quantile is estimated by $\widehat{VaR}_{q,t} = \sqrt{\widehat{V}_t}\,\hat{z}_{q,t}$. Here $\hat{z}_{q,t}$ is estimated from past returns standardized by the corresponding past variance forecasts. Only observations available before the target day enter the quantile estimate (Jorion, 2001).

The allocation exercise asks whether changing the variance forecast improves realized risk and net utility under the same exposure rule. The VaR exercise asks whether its predicted tail threshold has appropriate coverage and useful quantile accuracy. Neither exercise identifies a general economic value independently of its rule and sample.

A VaR violation occurs when the return falls below its forecast quantile. Kupiec's test checks the violation frequency against q, and Christoffersen's test adds a first-order independence assessment (Kupiec, 1995; Christoffersen, 1998). Quantile loss compares forecasts at a common tail level (Koenker and Bassett, 1978); sparse violations limit the precision of the coverage evidence.

Summary of the Chapter

The chapter leaves three things for the empirical application. The first is a forecasting target and an event set: the realized variance of a day, its continuous and jump components, and the returns that the intraday tests flag. The second is a model of those events with its tools of inference: the Hawkes process, its maximum likelihood estimate, a simulation algorithm and likelihood ratio tests between nested processes. The third is the route from the fitted process to a forecast: the expected count of the next day in closed form, the regression in which it enters beside the alternatives, and the tests that compare the forecasts. Chapter 3 takes them in that order: it compares the jump regressors, then sets process fit and all-return controls against the forecasting results. Simulation diagnostics and the S&P 500 application are documented in the appendices.

Chapter 3

Empirical Application

This chapter asks whether intraday jump information improves Bitcoin volatility forecasts, whether richer event-process models improve those forecasts, and whether the gains remain beside recent volatility measured from all returns. Data and full-sample fits precede the matched forecast comparisons, followed by all-return controls, six design sensitivities and an economic exercise. Ethereum and Litecoin assess transfer to related assets. Simulation checks and the S&P 500 application are collected in Appendices B and C.

3.1 Data Setup and In-Sample Analysis

3.1.1 Data Sources and Return Construction

Bitcoin, Ethereum and Litecoin are observed through Binance Spot USDT pairs using monthly exchange klines (Binance, 2026). Five-minute bars give 288 intervals per UTC calendar day. Bitcoin and Ethereum span 17 August 2017 to 31 July 2026; Litecoin starts on 13 December 2017. These observations represent prices on one exchange, and the two additional cryptocurrencies provide related-asset comparisons.

A day is retained when at least 80% of its nominal bars are stored. For opening price $O_{t,1}$ and bar closes $P_{t,i}$, percentage log returns are $r_{t,1} = 100 \log(P_{t,1}/O_{t,1})$ and $r_{t,i} = 100 \log(P_{t,i}/P_{t,i-1})$ for $i \geq 2$. A missing bar takes the last observed price and therefore gives a zero return; a stored bar with an unchanged close remains a genuine zero return. Realized variance is consequently measured in squared percentage points.

Table 3.1 separates missing observations from unchanged prices. Bitcoin and Ethereum retain 3,261 days and Litecoin 3,144; fewer than 0.8% of retained

cryptocurrency days fall below 99% bar coverage. No duplicate or off-grid timestamp is stored. Missing bars account for 0.067% of the five-minute cells on each cryptocurrency, whereas genuine zero returns are more common, particularly on Litecoin, where 5.17% of returns are zero. That difference limits direct interpretation of detector comparisons across assets.

Table 3.1: The stored price series and their coverage

	BTC	ETH	LTC	SPX
Venue	Binance	Binance	Binance	Index
Series	BTC/USDT	ETH/USDT	LTC/USDT	S&P 500
First day	2017-08-17	2017-08-17	2017-12-13	2004-01-02
Last day	2026-07-31	2026-07-31	2026-07-31	2019-10-18
Stored days	3,271	3,271	3,153	3,977
Retained days	3,261	3,261	3,144	3,944
Regression days	2,981	2,981	2,864	—
Bars a day	288	288	288	78
Missing bars (%)	0.067	0.067	0.067	0.015
Days with a missing bar	25	25	24	16
Days below 99% coverage (%)	0.77	0.77	0.76	0.41
Days below 95% coverage (%)	0.52	0.52	0.54	0.08
Zero returns (%)	0.54	1.16	5.17	0.56
Days with $\geq$5% zero returns	82	126	1,132	15
Forecast origins	1,981	1,981	1,864	2,664

Five-minute bars; S&P 500 days are trading sessions. A day is retained when at least 80% of its bars are stored; first and last day refer to retained days. Regression days start at the 281st retained day, the first with a full 30-day cascade. Missing bars: share of the five-minute cells of the retained days with no stored bar; a missing bar takes the last price. Coverage: share of a day's bars that are stored. No duplicate or off-grid timestamp is stored. Zero-return days: days on which at least 5% of the five-minute returns are zero. Forecast origins: one-day forecasts of the main design.

The S&P 500 column gives the coverage of the supplementary equity sample; its session clock, return construction and forecast results are developed together in Appendix C. Data handling and computational settings are collected in Appendix E. All comparisons below use the observations available to each asset, so differences in absolute losses across assets are not interpreted as differences in model quality.

3.1.2 Realized Measures and Jump Definitions

The primary design uses five-minute realized measures, the bipower HAR-CJ benchmark of Section 2.4.1, and the Intersection of the ABD, Lee–Mykland and sequential BNS event sets. The intraday tests use a 1% daily level; the BNS test also uses 1% and identifies jump days for the daily process and the post-jump

forecast subset. The Intersection requires agreement among detectors and is deliberately sparse. Alternative detectors and sampling grids are examined in Section 3.5.

The Bitcoin measures in Table 3.2 describe 2,981 regression days. Log realized variance has mean 1.820 and standard deviation 1.073, while its kurtosis of 3.723 indicates heavier tails than a Gaussian distribution. The test-based jump measure is zero on 68% of days; the positive part of $RV - BPV$, used in HAR-CJ, is zero on only 16.5%. The transformation $\log(1 + J)$ retains those zero observations without equating the two definitions of jump variation.

Table 3.2: Distribution of the daily realized measures: Bitcoin

Measure	n	Mean	Median	SD	Skewness	Kurtosis	$q_{0.01}$	$q_{0.05}$	$q_{0.95}$	$q_{0.99}$	Zeros (%)
$\log RV$	2,981	1.820	1.810	1.073	0.135	3.723	−0.697	0.080	3.635	4.522	—
$\log BPV$	2,981	1.698	1.692	1.064	0.119	3.876	−0.792	−0.056	3.471	4.334	—
$\log C^{BNS}$	2,981	1.728	1.716	1.064	0.124	3.843	−0.769	−0.024	3.506	4.349	—
$\log(1 + J^{BNS})$	2,981	0.347	0.000	0.664	2.274	8.385	0.000	0.000	1.870	2.714	68.0
$\log(1 + J)$, $J = \max(RV - BPV, 0)$	2,981	0.562	0.323	0.666	1.840	6.990	0.000	0.000	1.943	2.789	16.5

Bitcoin, daily measures from five-minute percentage log returns, regression sample 2018-05-27 to 2026-07-31. RV and BPV in squared percent. C^{BNS} and J^{BNS}: continuous and jump variation separated by the daily BNS test at 1% ($C^{BNS} = BPV$ and $J^{BNS} = RV - BPV$ on a jump day; $C^{BNS} = RV$ and $J^{BNS} = 0$ otherwise). The jump measures enter as $\log(1 + \cdot)$ so that zero days remain defined; Zeros is their share of zero days. Kurtosis is not in excess.

Table 3.3 shows the reduction in event density from requiring agreement. Bitcoin has 1,382 Intersection events over the regression sample, about 0.46 per day, compared with 1.69 for ABD, 1.90 for Lee–Mykland and 1.17 for sequential BNS. Their Union reaches 2.82 events per day. Ethereum and Litecoin have similarly sparse Intersection sets, but these counts measure detector output and do not establish which set contains the true price jumps.

3.1.3 In-Sample Analysis

Full-sample maximum likelihood estimates describe the event processes before their forecasting role is assessed. Table 3.4 uses one parameter schema for the unmarked, marked and two-exponential specifications, with Hessian-based standard errors in log parameters and the delta method. Twenty random starts confirmed the reported optima; this numerical check does not prove global optimality. Simulation evidence on recovery and finite-sample precision is given in Appendix B.

The one-exponential intraday fits have branching ratios of 0.17–0.20 and half-lives of 26–36 minutes. Adding a second component changes that description: Bitcoin combines a 22-minute component with a 9.6-day component, and its

Table 3.3: Jump days and intraday jump events by asset and detector

Asset	Detector	Jump days	Events	Events a day	Days with an event (%)
Bitcoin	BNS (daily)	954	—	—	32.0
	Intersection	—	1,382	0.46	28.3
	ABD	—	5,031	1.69	80.3
	Lee–Mykland	—	5,650	1.90	71.8
	s-BNS	—	3,494	1.17	32.0
	Union	—	8,397	2.82	84.4
Ethereum	BNS (daily)	1,045	—	—	35.1
	Intersection	—	1,536	0.52	31.9
	ABD	—	5,088	1.71	80.7
	Lee–Mykland	—	5,480	1.84	72.9
	s-BNS	—	4,230	1.42	35.1
Litecoin	BNS (daily)	931	—	—	32.5
	Intersection	—	1,248	0.44	28.0
	ABD	—	4,331	1.51	75.5
	Lee–Mykland	—	4,541	1.59	67.4
	s-BNS	—	4,722	1.65	32.5

Regression sample of each asset (2,981 days for Bitcoin and Ethereum, 2,864 for Litecoin); every rate uses that number of days. BNS: daily ratio test at 1%. Intraday detectors on five-minute returns: ABD, Lee–Mykland and sequential BNS (s-BNS); the Intersection keeps the returns flagged by all three, the Union those flagged by any. —: not defined for the detector.

branching ratio rises from 0.176 to 0.794. The daily process instead has a 17.4-day half-life; Ethereum and Litecoin also have daily memory measured in weeks. These are model-dependent descriptions of clustering, and a branching ratio estimated under one kernel is not an invariant property of the market.

The intraday mark contributes little to the full-sample fits in Table 3.4. Bitcoin's one-exponential exponent is 0.031 with standard error 0.056, and Litecoin's two-exponential exponent reaches zero. The evidence therefore does not support treating every marked estimate as an interior parameter with a precise effect. For marked processes the reported branching ratio averages impact over observed marks; interpreting it as expected offspring requires the conditional mark-moment assumption of Chapter 2.

The daily and weekly continuous components account for most of the persistence in the full-sample Bitcoin HAR-CJ regression (Table 3.5). Their coefficients are 0.430 and 0.387, while the monthly continuous term is not significant. Among the three jump terms, the monthly term is marginal at 10%, with a t statistic of 1.80. Thus the regression provides weak evidence for jump magnitudes, rather than a complete absence of association.

EI, LCI, F and DI enter positively in Table 3.5, while SI is not significant. Their contribution to fit is small: adjusted R^2 rises from 0.551 to at most 0.554.

Table 3.4: Full-sample Hawkes fits of the model ladder

Asset	Process	Events	μ	α_1	β_1	α_2	β_2	δ	Log-likelihood	Half-lives	η	Optimum
Bitcoin	One exponential	1,532	0.387	0.176	30.532	—	—	—	−2344.7	33 min	0.176	confirmed
			(0.012)	(0.015)	(5.755)							
	Two exponentials	1,532	0.098	0.148	46.053	0.646	0.072	—	−2214.2	22 min; 9.6 d	0.794	confirmed
			(0.022)	(0.012)	(6.713)	(0.049)	(0.022)					
	One exponential, mark	1,532	0.388	0.198	31.720	—	—	0.031	−2344.5	31 min	0.173	confirmed
			(0.012)	(0.046)	(6.237)			(0.056)				
	Two exponentials, mark	1,532	0.094	0.184	46.142	0.813	0.065	0.051	−2213.3	22 min; 10.7 d	0.803	confirmed
			(0.020)	(0.033)	(6.583)	(0.146)	(0.018)	(0.038)				
	Daily	1,090	0.096	0.720	0.040	—	—	—	−2227.2	17.4 d	0.720	confirmed
			(0.021)	(0.068)	(0.007)							
	Daily, MedRV mark	1,090	0.103	1.214	0.037	—	—	0.142	−2223.7	18.5 d	0.698	confirmed
			(0.022)	(0.256)	(0.007)			(0.058)				
Ethereum	One exponential	1,673	0.412	0.198	28.100	—	—	—	−2353.6	36 min	0.198	confirmed
			(0.012)	(0.014)	(3.645)							
	Two exponentials	1,673	0.117	0.166	40.831	0.608	0.083	—	−2229.9	24 min; 8.3 d	0.774	confirmed
			(0.019)	(0.013)	(5.501)	(0.041)	(0.014)					
	One exponential, mark	1,673	0.412	0.242	28.723	—	—	0.051	−2353.2	35 min	0.196	confirmed
			(0.012)	(0.053)	(3.756)			(0.053)				
	Two exponentials, mark	1,673	0.117	0.189	40.985	0.694	0.082	0.032	−2229.5	24 min; 8.5 d	0.776	confirmed
			(0.019)	(0.033)	(5.499)	(0.122)	(0.014)	(0.039)				
	Daily	1,154	0.117	0.674	0.049	—	—	—	−2301.9	14.1 d	0.674	confirmed
			(0.023)	(0.067)	(0.008)							
	Daily, MedRV mark	1,154	0.118	0.982	0.048	—	—	0.092	−2300.6	14.4 d	0.671	confirmed
			(0.023)	(0.241)	(0.008)			(0.059)				
Litecoin	One exponential	1,409	0.373	0.168	37.755	—	—	—	−2196.9	26 min	0.168	confirmed
			(0.011)	(0.014)	(5.930)							
	Two exponentials	1,409	0.071	0.144	53.842	0.700	0.049	—	−2056.2	19 min; 14.0 d	0.844	confirmed
			(0.017)	(0.012)	(7.991)	(0.043)	(0.010)					
	One exponential, mark	1,409	0.373	0.214	38.334	—	—	0.035	−2196.7	26 min	0.167	confirmed
			(0.012)	(0.092)	(6.093)			(0.061)				
	Two exponentials, mark	1,409	0.071	0.144	53.842	0.700	0.049	0.000	−2056.2	19 min; 14.0 d	0.844	δ at 0
			(0.017)	(0.012)	(7.991)	(0.043)	(0.010)	(bound)				
	Daily	1,043	0.067	0.807	0.026	—	—	—	−2122.8	26.2 d	0.807	confirmed
			(0.019)	(0.063)	(0.005)							
	Daily, MedRV mark	1,043	0.067	0.884	0.026	—	—	0.019	−2122.7	26.5 d	0.809	confirmed
			(0.019)	(0.330)	(0.005)			(0.076)				

Full-sample maximum likelihood. Intraday processes on the Intersection events, daily ones on the BNS jump days; time in days, so μ is a rate a day and β a decay a day. One-exponential processes use the first channel; a mark enters as m^{δ}, with m the squared return (intraday) or the day's MedRV (daily) divided by its sample maximum. Standard errors in parentheses, from the Hessian of the log-likelihood in log-parameters (delta method). η: branching ratio, the mean impact over the observed marks for marked processes. Optimum: confirmed by 20 random starts with a definite Hessian; "δ at 0" marks a mark exponent on its bound. Log-likelihoods compare only within one asset and event set.

Table 3.5: Full-sample fit of HAR-CJ and its extensions: Bitcoin

Model	Const.	Continuous component			Jump component			Added regressor	Adj. R^2	Log-likelihood
		C_d	C_w	C_m	J_d	J_w	J_m			
HAR-CJ	0.277*** (8.12)	0.430*** (18.92)	0.387*** (9.85)	0.038 (0.83)	0.033 (1.29)	−0.005 (−0.09)	0.135* (1.80)		0.551	-3243.6
HAR-CJ-DI	0.115* (1.75)	0.430*** (19.01)	0.396*** (10.12)	0.094** (2.01)	0.032 (1.24)	−0.021 (−0.33)	−0.050 (−0.50)	0.497** (2.52)	0.551	-3241.5
HAR-CJ-SI	0.231*** (3.69)	0.435*** (18.37)	0.388*** (9.78)	0.037 (0.82)	0.018 (0.52)	−0.007 (−0.11)	0.132* (1.78)	0.101 (0.88)	0.551	-3243.2
HAR-CJ-EI	−0.501** (−2.24)	0.439*** (19.17)	0.384*** (9.94)	0.040 (0.92)	0.016 (0.64)	0.001 (0.01)	0.122 (1.64)	1.645*** (3.50)	0.552	-3237.3
HAR-CJ-LCI	1.043*** (5.05)	0.440*** (19.21)	0.383*** (9.93)	0.040 (0.92)	0.014 (0.55)	0.001 (0.01)	0.122 (1.64)	1.020*** (3.78)	0.553	-3236.5
HAR-CJ-F	0.250*** (7.52)	0.451*** (19.20)	0.381*** (9.85)	0.044 (0.99)	−0.014 (−0.52)	−0.003 (−0.06)	0.124* (1.68)	0.263*** (5.06)	0.554	-3231.8
HAR-CJ-EWRV	1.357*** (16.54)	0.223*** (8.74)	0.314*** (8.50)	0.044 (1.08)	−0.050** (−2.08)	0.024 (0.39)	0.049 (0.68)	0.297*** (14.13)	0.592	-3098.2

OLS of $\log RV$ over the full Bitcoin sample (2,981 days), HAR-CJ with the 1/7/30-day cascade; C is bipower variation and $J = \max(RV - BPV, 0)$. Newey–West t-statistics (30 lags) in parentheses. Gaussian log-likelihood of the OLS fit. These are in-sample associations: forecast performance is assessed out of sample. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

EWRV raises it to 0.592 and reduces the coefficient on the daily continuous term by about half. These full-sample associations motivate the forecast comparisons; they do not measure out-of-sample accuracy. The HAR-RV regressions are retained in Appendix D.1.

3.2 Out-of-Sample Forecasting

3.2.1 Forecast Design and Evaluation

Every forecast uses information available through the previous close. The primary design is a rolling 1,000-day window, five-minute Intersection events, the bipower HAR-CJ benchmark with a 1/7/30-day cascade, and daily refits. Table 3.6 gives 1,981 Bitcoin and Ethereum forecasts and 1,864 Litecoin forecasts; 541, 632 and 532 targets, respectively, follow a BNS jump day. Each comparison is evaluated on common origins within an asset.

Table 3.6: Forecast samples by asset and refit schedule

Asset	Process refit	First target	Last target	Forecasts	After a jump day
Bitcoin	Daily	2021-02-27	2026-07-31	1,981	541
	30 days	2021-02-27	2026-07-31	1,981	541
Ethereum	Daily	2021-02-27	2026-07-31	1,981	632
Litecoin	Daily	2021-06-24	2026-07-31	1,864	532

One-day-ahead forecasts of $\log RV$ from a rolling window of 1,000 days. Common to every row: Intersection events on five-minute returns, bipower HAR-CJ with the 1/7/30-day cascade, and HAR coefficients re-estimated every day. The smallest Hawkes estimation window holds 220 events. The process refit is the interval between Hawkes refits. After a jump day: forecasts whose preceding day is a BNS jump day.

At each origin the history of each jump regressor is recomputed using that origin's parameter estimate, so the affine equivalence of Section 2.3.1 holds within the regression. EI is implemented through the equivalent conditional expected count. The 30-day design holds process parameters and tuned decay rates fixed between refits while updating states with observed events; HAR coefficients are re-estimated daily. This schedule provides matched comparisons of richer processes and design sensitivities. Appendix E records initialization and eligibility rules.

MSE and MAE evaluate log realized variance; QLIKE and HRMSE evaluate variance levels after the lognormal correction of Section 2.5.2. Paired DM–HLN tests use Holm adjustment over the four losses of a model pair unless a table states another family. The six-model confidence set contains HAR-CJ and its

DI, SI, EI, LCI and F extensions. It uses T_R and $T_{\max}$, moving blocks of 12 days, 10,000 draws for the two main panels, and retention at adjusted $p > 0.25$. Comparisons across overlapping configurations remain descriptive.

3.2.2 Comparison of Jump-Arrival Regressors

DI summarizes the daily process and SI integrates intraday intensity over the completed day. EI retains the state at the close, LCI transforms the conditional count logarithmically, and F chooses the memory of a decayed count by forecast loss. Each is added separately to the same HAR-CJ benchmark, so Table 3.7 compares alternative summaries of jump arrivals without changing the volatility regression.

Table 3.7: Primary intraday comparison: Bitcoin

	Overall				Following jump days			
Model	QLIKE	MSE	MAE	HRMSE	QLIKE	MSE	MAE	HRMSE
HAR-CJ	0.3210	0.5216	0.5303	3.2518	0.2980	0.4924	0.5245	2.0657
DI	+0.06	−0.06	−0.09	−0.48	+0.69	+0.27	−0.06	−0.10
SI	+0.13	+0.06	+0.07	−0.21	+0.96*	+0.66	+0.30	+2.90*
EI	−0.06	−1.16**	−0.61	−4.67**	−1.76	−3.50*	−1.94	−3.87
LCI	−0.12	−1.30**	−0.62	−5.11**	−1.98	−3.93**	−1.96	−3.86
F	+0.05	−0.99*	−0.22	−4.46**	−1.36	−3.26*	−1.48	−2.25

Benchmark rows give absolute mean losses; all other rows give percentage changes against their matching benchmark (negative favors the extension). * and **: two-sided DM–HLN tests, Holm-adjusted, at 5% and 1%. Holm family: four losses for each pair and sample. $n = 1,981$ overall, 541 following BNS jump days. Five-minute intersection events, bipower HAR-CJ, 1/7/30 cascade, rolling 1,000-day estimation and daily process, half-life and regression refits. Each extension adds the indicated regressor to HAR-CJ.

On Bitcoin, EI lowers MSE by 1.16% and HRMSE by 4.67%, both significant at 1% after adjustment. LCI and F also improve these losses, while DI and SI show no significant gain under any of the four losses. None of EI, LCI and F significantly improves QLIKE or MAE. The EI MSE reduction is larger following a jump day, at 3.50%, so the main gain is concentrated in a subset of the forecast period.

The confidence sets in Table 3.8 retain EI, LCI and F under MSE with daily refits, and EI and LCI with 30-day refits. Both statistics retain only LCI under HRMSE and all six models under QLIKE. MAE membership depends on the statistic. These sets support the competitiveness of the closing-state regressors in the main design without selecting one model for every loss.

Table 3.8: Main-configuration model confidence set p-values with statistic-specific retention prevalence.

Panel A: Daily refits

	QLIKE		MSE		MAE		HRMSE	
Model	T_R	$T_{\max}$	T_R	$T_{\max}$	T_R	$T_{\max}$	T_R	$T_{\max}$
HAR-CJ	**0.959** (66.7%)	**0.990** (58.3%)	0.005 (8.3%)	0.006 (25.0%)	0.109 (33.3%)	0.131 (33.3%)	0.000 (0.0%)	0.000 (8.3%)
+ DI	**0.959** (58.3%)	**0.990** (66.7%)	0.020 (25.0%)	0.025 (33.3%)	0.195 (41.7%)	**0.377** (41.7%)	0.002 (0.0%)	0.001 (16.7%)
+ SI	**0.893** (66.7%)	**0.948** (66.7%)	0.005 (25.0%)	0.006 (33.3%)	0.109 (33.3%)	0.089 (33.3%)	0.000 (0.0%)	0.000 (16.7%)
+ EI	**0.948** (66.7%)	**0.990** (75.0%)	**0.265** (50.0%)	**0.319** (66.7%)	**0.911** (75.0%)	**0.911** (**83.3**%)	0.024 (8.3%)	0.188 (41.7%)
+ LCI	**1.000** (66.7%)	**1.000** (75.0%)	**1.000** (66.7%)	**1.000** (66.7%)	**1.000** (**91.7**%)	**1.000** (**83.3**%)	**1.000** (66.7%)	**1.000** (75.0%)
+ F	**0.948** (**100.0**%)	**0.990** (**100.0**%)	**0.265** (**83.3**%)	**0.319** (**91.7**%)	0.109 (58.3%)	**0.377** (75.0%)	0.024 (75.0%)	0.188 (75.0%)

Panel B: 30-day refits

	QLIKE		MSE		MAE		HRMSE	
Model	T_R	$T_{\max}$	T_R	$T_{\max}$	T_R	$T_{\max}$	T_R	$T_{\max}$
HAR-CJ	**0.939**	**0.988**	0.006	0.008	0.133	0.152	0.000	0.000
+ DI	**0.939**	**0.965**	0.024	0.037	**0.252**	**0.440**	0.002	0.002
+ SI	**0.887**	**0.965**	0.006	0.008	0.133	0.103	0.000	0.000
+ EI	**0.939**	**0.988**	**0.273**	**0.273**	**0.970**	**0.970**	0.017	0.032
+ LCI	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
+ F	**0.887**	**0.965**	0.131	0.162	0.133	**0.440**	0.007	0.032

Bitcoin, Intersection events, HAR-CJ, 1,981 matched forecast origins. Both panels use 10,000 moving-block draws, blocks of 12 days and the same seed. Panel B holds Hawkes fits and tuning for 30 days; OLS is re-estimated at every origin. Bold p-values exceed 0.25. Parentheses in Panel A only: percentage retained across twelve other configurations with 30-day refits and 2,000 draws, separately by statistic and loss; only percentages above 75% are bold. Backbone rows use each benchmark. T_R and $T_{\max}$ use their respective standardized elimination rules. Prevalence is descriptive, not a probability of superiority. MSE and MAE use log RV; QLIKE and HRMSE use level forecasts.

Across twelve other configurations, EI is retained under MSE in 50% of the range-statistic sets and 67% of the maximum-statistic sets, with lower retention under HRMSE. LCI and F are retained more often under these losses. Because the configurations share data, these percentages describe sensitivity to design choices. They are not independent replications or evidence that jointly retained forecasts have equal accuracy (Tables 3.8 and D.3).

EI and F share the decayed-count form of Section 2.3.2; they differ in how the decay is selected. The fixed-memory profile in Figure 3.1 reaches its Bitcoin MSE minimum around two hours. The rolling likelihood selects a median half-life of 98 minutes and F selects 130 minutes, leaving little separation in their main-design losses. The 98-minute rolling estimate differs from the 33-minute full-sample estimate because it is fitted on a changing window. The fixed-memory profile is descriptive and is not used to select a decay with knowledge of future outcomes.

3.2.3 Ethereum and Litecoin

The Bitcoin MSE gain does not extend significantly to Ethereum or Litecoin (Table 3.9). EI lowers HRMSE significantly on Ethereum, by 1.78%, and LCI lowers it on both assets, by 2.14% and 2.51%. The result is therefore loss-specific even among related markets. The supplementary equity application likewise

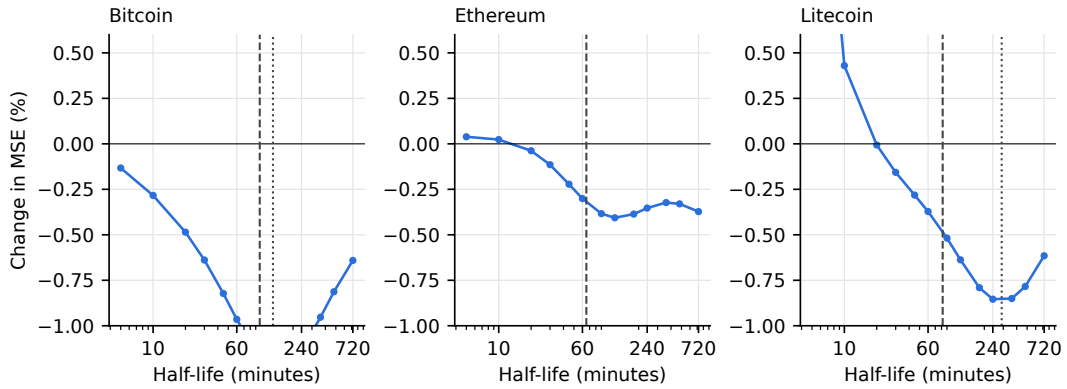


Figure 3.1: Loss of the decayed count as a function of its half-life. The count enters HAR-CJ at a half-life held fixed over the whole forecast period; each point is the percentage change in MSE against HAR-CJ under the primary daily-refit design. Dashed line: the median half-life of the daily likelihood fits of the intraday process. Dotted line: the median half-life of the daily tuning of F, drawn where it falls inside the range of the axis. The vertical axis is cut at 0.6; the five-minute point of Bitcoin and the five- and ten-minute points of Ethereum lie above it.

has a different pattern: the daily intensity improves HRMSE, but no extension improves the primary benchmark under MSE, MAE or QLIKE (Appendix C).

3.3 Fit and Forecast

A process can describe arrival clustering better without supplying a better regressor for realized variance. This section first tests the one-exponential process against homogeneous Poisson arrivals, then compares richer processes through both likelihood ratios and forecast losses. The two comparisons answer different parts of the research question.

3.3.1 Self-Excitation against a Poisson Process

The Poisson comparison uses each of Bitcoin's 1,981 rolling 1,000-day windows and a parametric-bootstrap critical value (Table 3.10). The decay is unidentified under the Poisson null, so the boundary-mixture value of 2.71, also shown for comparison, is not treated as its valid reference law. The calibration and its limits are described in Appendix B.9.

Poisson is rejected in every intraday Intersection and ABD window. On BNS jump days it is rejected in 86.2% of windows, compared with only 22.8% for days containing an ABD flag. The latter series is active on most days and frequently

Table 3.9: Primary intraday comparison: Ethereum and Litecoin

Asset	Model	QLIKE	MSE	MAE	HRMSE
Ethereum	HAR-CJ	0.2842	0.4369	0.4873	1.7368
	DI	−0.12	−0.01	−0.01	−0.71
	SI	−0.33	−0.02	−0.00	+0.39
	EI	+0.25	−0.28	−0.16	−1.78**
	LCI	+0.20	−0.39	−0.26	−2.14**
	F	+0.31	−0.31	−0.33	−0.08
Litecoin	HAR-CJ	0.3291	0.4048	0.4710	1.0396
	DI	+0.09	+0.13	+0.18	−0.26
	SI	+0.10	−0.10	+0.02	−1.30
	EI	−0.63	−0.48	−0.21	−1.83
	LCI	−0.93	−0.71	−0.30	−2.51**
	F	−0.44	−0.09	+0.09	−2.39

HAR-CJ rows give absolute mean losses; the other rows give percentage changes against HAR-CJ (negative favors the extension). * and **: significant at 5% and 1% after the Holm adjustment. Daily refits, five-minute Intersection events, 1/7/30 cascade, 1,981 and 1,864 forecasts.

Table 3.10: Self-excitation against a Poisson process on rolling windows: Bitcoin

Event set	Time scale	Events (median)	Likelihood ratio		Critical value	Poisson rejected (%)		Boundary fits
			median	interquartile	bootstrap 5%	bootstrap	> 2.71	
Intersection	intraday	501	220.73	190.13–254.13	5.17	100.0	100.0	0
ABD	intraday	1,693	671.22	612.51–714.71	5.18	100.0	100.0	0
Intersection days	daily	295	11.97	7.46–18.19	1.98	94.5	93.8	145
BNS days	daily	337	7.14	3.53–17.58	2.00	86.2	81.2	120
ABD days	daily	805	0.00	0.00–0.51	0.62	22.8	13.7	471

Bitcoin, one-exponential Hawkes process against a homogeneous Poisson process on the rolling 1,000-day window of each of the 1,981 forecast origins. Likelihood ratio: twice the log-likelihood gain. The critical value is the 95% quantile of the parametric bootstrap under the fitted Poisson null; 2.71 is the 5% value of the boundary mixture $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$. Boundary fits: windows whose fitted excitation is at the bound. Daily sets treat each jump day as one event.

reaches the excitation boundary. Boundary fits in this table are parameter estimates at the constraint, not counts of numerical failures. The rolling Poisson test was not computed for Ethereum, Litecoin or the S&P 500.

Likelihood support for self-excitation does not establish that the Hawkes model describes all features of the arrivals. The grid-aware full-sample Bitcoin diagnostics reject both intraday kernels under the distance statistics, although the second component removes the detected duration dependence. Those diagnostics are reported in Appendix B.10. Forecast losses therefore remain a separate criterion for choosing a regressor.

3.3.2 Richer Processes and Forecast Accuracy

Table 3.11 pairs each richer process with its nested simpler process on identical origins, using 30-day refits. It compares two exponentials against one, a power mark against no mark, a mark on two exponentials, and a mark on the daily process. Each loss change is relative to the regressor of that row's simpler process; a daily marked gain is consequently a gain over DI, not a direct comparison with EI.

Table 3.11: Fit against forecast: richer Hawkes processes and their regressors

		Likelihood ratio					Loss change (%)			
Asset	Richer | simpler process	full sample	critical	rolling rejected (%)	windows	Forecasts	QLIKE	MSE	MAE	HRMSE
Bitcoin	Two exponentials | one	260.95	5.00	100	67	1,981	+0.40	+0.46	+0.22	+0.40
	Marks | none	0.30	2.71	21	67	1,981	+0.01	+0.01	−0.00	−0.01
	Marks | two exponentials	1.87	2.71	0	67	1,981	+0.01	−0.00	−0.00	−0.00
	Daily: marks | none	6.88	2.71	69	67	1,981	−1.07	−0.28	−0.12	−2.69*
Ethereum	Two exponentials | one	247.59	5.30	100	67	1,981	+1.17	+0.12	−0.03	+1.14
	Marks | none	0.93	2.71	0	67	1,981	−0.00	−0.01	−0.02	−0.04*
	Marks | two exponentials	0.65	2.71	3	67	1,981	−0.00	+0.02	+0.00	+0.06
	Daily: marks | none	2.52	2.71	43	67	1,981	+0.12	+0.13	−0.25	−1.09
Litecoin	Two exponentials | one	281.42	4.52	100	63	1,864	+0.64	+0.28	+0.20	+0.18
	Marks | none	0.33	2.71	0	63	1,864	−0.01	−0.01	−0.01	−0.02
	Marks | two exponentials	0.00	2.71	0	63	1,864	+0.01	+0.00	−0.00	−0.02
	Daily: marks | none	0.06	2.71	6	63	1,864	+0.10	+0.08	+0.01	−0.64**

Each row nests the simpler process in the richer one, on Intersection events (intraday) or BNS jump days (daily). Likelihood ratio: full sample and the share of rolling 1,000-day windows, refitted every 30 days, that reject the simpler process at 5%; the critical value is the parametric-bootstrap quantile for two exponentials and the boundary mixture 2.71 for a mark. Loss change: the richer regressor against the simpler one in HAR-CJ, on the same one-day forecasts with fits held for 30 days; negative favors the richer. * and **: two-sided DM–HLN tests, Holm-adjusted over the four losses of a pair, at 5% and 1%.

The second exponential is preferred at every recorded refit on all three cryptocurrencies, but none of its four whole-sample loss changes is significant. On Bitcoin it increases MSE by 0.46%. Re-estimating both processes daily also gives no significant whole-sample gain over EI; alternative state summaries are

reported in Appendix D.3. The evidence separates a better fit to arrival times from improved variance prediction.

Intraday marking gives very small loss changes in Table 3.11. The one-exponential mark reduces Ethereum HRMSE by 0.04%, significantly at 5%, while the other whole-sample changes are not significant. Adding a mark to the two-exponential process yields no significant improvement on any of the three assets. The exception prevents a blanket claim that richer intraday specifications never improve a loss, but it provides little evidence of a broad predictive benefit.

Daily marking has a different result: it lowers DI's HRMSE by 2.69% on Bitcoin and 0.64% on Litecoin, with adjusted significance at 5% and 1%, respectively. Its QLIKE, MSE and MAE changes are not significant. These are results for the 30-day panel. A complete daily-refit marked ladder was not computed, and the reported daily-refit two-exponential check covers Bitcoin only.

3.4 The All-Return Controls

HAR-CJ contains daily aggregates. An intraday arrival regressor may improve it by transmitting recent volatility that those aggregates omit. The following comparisons add two measures constructed from all returns and test whether EI supplies a further gain.

3.4.1 The Two Controls

RV60 sums the squared returns of the last hour. EWRV discounts all past squared returns with a half-life selected on the training window (Section 2.3.2). Each enters HAR-CJ as a separate control, and EI is then added to the controlled regression. Neither control requires an event detector.

Both controls improve every loss on Bitcoin (Table 3.12). RV60 reduces MSE by 11.85% and EWRV by 12.93%, compared with 1.16% for EI. Adding EI reduces their MSE by a further 0.14% and 0.16%, respectively, but no increment is significant under any of the four losses. The benchmark therefore determines whether the intensity has an established accuracy gain.

The comparison carries over to Ethereum and Litecoin in Table 3.13. Both controls reduce MSE substantially, while adding EI to either produces no significant whole-sample gain under the reported losses. These results contrast with the limited transfer of the uncontrolled Bitcoin jump-regressor gain.

Table 3.12: The all-return controls and the jump regressor: Bitcoin

Model	QLIKE	MSE	MAE	HRMSE
HAR-CJ (mean loss)	0.3210	0.5216	0.5303	3.2518
EI	−0.06	−1.16**	−0.61	−4.67**
RV60	−11.36**	−11.85**	−4.69**	−34.06**
EWRV	−11.53**	−12.93**	−5.39**	−37.97**
RV60 + EI	−11.18**	−11.97**	−4.81**	−34.50**
EWRV + EI	−11.57**	−13.08**	−5.53**	−38.19**
EI added to RV60	+0.20	−0.14	−0.12	−0.66
EI added to EWRV	−0.04	−0.16	−0.15	−0.36

First row: mean loss of HAR-CJ. Other rows: percentage change in mean loss against HAR-CJ or, in the last two rows, against the control alone (negative favors the first model). Identical one-day forecast origins within an asset (Bitcoin 1,981). RV60: realized variance of the last hour; EWRV: exponentially weighted realized variance. * and **: two-sided DM–HLN tests, Holm-adjusted over the four losses of a pair, at 5% and 1%.

Table 3.13: The all-return controls and the jump regressor: Ethereum and Litecoin

Asset	Model	QLIKE	MSE	MAE	HRMSE
Ethereum	HAR-CJ (mean loss)	0.2842	0.4369	0.4873	1.7368
	EI	+0.25	−0.28	−0.16	−1.78**
	RV60	−6.63**	−8.41**	−3.93**	−27.13**
	EWRV	−6.80*	−8.87**	−4.28**	−28.78**
	RV60 + EI	−6.90**	−8.49**	−3.97**	−27.13**
	EWRV + EI	−7.54*	−9.16**	−4.29**	−29.17**
	EI added to RV60	−0.28	−0.09	−0.04	−0.01
	EI added to EWRV	−0.79	−0.32	−0.01	−0.56
Litecoin	HAR-CJ (mean loss)	0.3291	0.4048	0.4710	1.0396
	EI	−0.63	−0.48	−0.21	−1.83
	RV60	−11.91	−8.34**	−3.72**	−18.35**
	EWRV	−11.63	−8.07**	−3.83**	−16.00**
	RV60 + EI	−11.76	−8.19**	−3.64**	−18.46**
	EWRV + EI	−11.44	−7.92**	−3.75**	−16.12**
	EI added to RV60	+0.18	+0.17	+0.08	−0.14
	EI added to EWRV	+0.21	+0.16	+0.09	−0.13

First row: mean loss of HAR-CJ. Other rows: percentage change in mean loss against HAR-CJ or, in the last two rows, against the control alone (negative favors the first model). Identical one-day forecast origins within an asset (Ethereum 1,981, Litecoin 1,864). RV60: realized variance of the last hour; EWRV: exponentially weighted realized variance. * and **: two-sided DM–HLN tests, Holm-adjusted over the four losses of a pair, at 5% and 1%.

3.4.2 Three Complementary Tests

The three tests examine different meanings of incremental information. The paired test compares achieved forecast losses; Clark–West adjusts the nested comparison for estimation noise; encompassing asks whether a second forecast contributes to a combination. Non-rejection by one test does not settle the other two questions.

Table 3.14 finds no significant paired gain from adding EI to the controls. Clark–West does reject on Ethereum when the benchmark contains EWRV, alone or together with RV60, with adjusted $p = 0.015$. The corresponding MSE reductions are 0.32% and 0.33%. This is evidence under the nested-model correction, while the achieved accuracy differences remain insignificant in the paired tests.

Table 3.14: Whether the jump regressor adds to an all-return control

		Loss change from adding EI (%)				Clark–West	
Asset	Control	QLIKE	MSE	MAE	HRMSE	statistic	Holm p
Bitcoin	none (HAR-CJ)	−0.06	−1.16**	−0.61	−4.67**	5.66	< 0.001
	RV60	+0.20	−0.14	−0.12	−0.66	1.97	0.127
	EWRV	−0.04	−0.16	−0.15	−0.36	2.05	0.127
	EWRV and RV60	−0.06	−0.17	−0.16	−0.31	2.08	0.127
Ethereum	none (HAR-CJ)	+0.25	−0.28	−0.16	−1.78**	3.31	0.005
	RV60	−0.28	−0.09	−0.04	−0.01	1.66	0.174
	EWRV	−0.79	−0.32	−0.01	−0.56	2.77	0.015
	EWRV and RV60	−0.79	−0.33	−0.09	−0.58	2.81	0.015
Litecoin	none (HAR-CJ)	−0.63	−0.48	−0.21	−1.83	3.84	< 0.001
	RV60	+0.18	+0.17	+0.08	−0.14	−0.76	1.000
	EWRV	+0.21	+0.16	+0.09	−0.13	−0.19	1.000
	EWRV and RV60	+0.21	+0.13	+0.08	−0.20	−0.03	1.000

Percentage change in mean loss when EI is added to the model with the control (negative favors adding EI), on identical one-day forecasts. * and **: two-sided DM–HLN tests, Holm-adjusted over the four losses of a pair, at 5% and 1%. Clark–West: one-sided test that the larger model's forecast adds to the nested one, on squared errors of $\log RV$, Holm-adjusted.

Table 3.15 does not reject that either control forecast encompasses EI, and rejects the reverse direction. The pattern holds in the cryptocurrency panels at both refit schedules and across the reported single-detector sets. Negative combination weights on EI are common, but not universal, and the tests do not prove the absence of all useful jump information. The S&P 500 rows are interpreted with the equity controls in Appendix C.6.

Table 3.15: Forecast encompassing of the jump regressor by the controls

					Weight on EI		Weight on control	
Event set	Process refit	Asset	Control	Forecasts	$\hat{\lambda}$	Holm p	$\hat{\lambda}$	Holm p
Intersection	daily refits	BTC	EWRV	1,981	−0.208	1.000	1.208	< 0.001
		BTC	RV60	1,981	−0.194	1.000	1.194	< 0.001
		ETH	EWRV	1,981	−0.136	1.000	1.136	< 0.001
		ETH	RV60	1,981	−0.202	1.000	1.202	< 0.001
		LTC	EWRV	1,864	−0.083	1.000	1.083	< 0.001
		LTC	RV60	1,864	−0.114	1.000	1.114	< 0.001
Intersection	30-day refits	BTC	EWRV	1,981	−0.208	1.000	1.208	< 0.001
		BTC	RV60	1,981	−0.196	1.000	1.196	< 0.001
		ETH	EWRV	1,981	−0.142	1.000	1.142	< 0.001
		ETH	RV60	1,981	−0.204	1.000	1.204	< 0.001
		LTC	EWRV	1,864	−0.079	1.000	1.079	< 0.001
		LTC	RV60	1,864	−0.111	1.000	1.111	< 0.001
ABD	30-day refits	BTC	EWRV	1,981	−0.201	1.000	1.201	< 0.001
		BTC	RV60	1,981	−0.222	1.000	1.222	< 0.001
		ETH	EWRV	1,981	−0.133	1.000	1.133	< 0.001
		ETH	RV60	1,981	−0.215	1.000	1.215	< 0.001
		LTC	EWRV	1,864	−0.060	1.000	1.060	< 0.001
		LTC	RV60	1,864	−0.110	1.000	1.110	< 0.001
Lee-Mykland	30-day refits	BTC	EWRV	1,981	−0.211	1.000	1.211	< 0.001
		BTC	RV60	1,981	−0.228	1.000	1.228	< 0.001
		ETH	EWRV	1,981	−0.141	1.000	1.141	< 0.001
		ETH	RV60	1,981	−0.225	1.000	1.225	< 0.001
		LTC	EWRV	1,864	−0.085	1.000	1.085	< 0.001
		LTC	RV60	1,864	−0.141	1.000	1.141	< 0.001
s-BNS	30-day refits	BTC	EWRV	1,981	−0.174	1.000	1.174	< 0.001
		BTC	RV60	1,981	−0.202	1.000	1.202	< 0.001
		ETH	EWRV	1,981	−0.105	1.000	1.105	< 0.001
		ETH	RV60	1,981	−0.165	1.000	1.165	< 0.001
		LTC	EWRV	1,864	0.028	1.000	0.972	< 0.001
		LTC	RV60	1,864	−0.013	1.000	1.013	< 0.001
Intersection	daily refits	SPX	EWRV	2,664	0.120	0.805	0.880	< 0.001
		SPX	RV60	2,664	0.013	0.932	0.987	< 0.001

Forecast-encompassing tests on the one-day $\log RV$ forecasts of the whole sample, HAR-CJ backbone; $\hat{\lambda}$ are the estimated combination weights. Weight on EI: H_0 that the control's forecast encompasses EI's ($\lambda = 0$); weight on control: H_0 that EI's forecast encompasses the control's. p-values are Holm-adjusted within each panel. SPX: the S&P 500.

3.4.3 The Last Hours of the Day

The memory profiles in Table 3.16 place the useful all-return information near the forecast origin. EWRV has its lowest reported MSE at 45 minutes on Bitcoin and Ethereum and at 30 minutes on Litecoin. Half-lives of a day or more change MSE by less than 0.4%, without significance. The trailing-window comparison also favors a small number of recent hours over the entire day. These fixed-memory profiles describe sensitivity and do not replace training-window tuning.

Table 3.16: The memory of the all-return control

Memory	Bitcoin	Ethereum	Litecoin
Panel A: EWRV, half-life			
5 min	−7.41**	−4.54**	−5.06**
15 min	−11.36**	−7.85**	−8.22**
30 min	−12.87**	−8.92**	−8.72**
45 min	−12.96**	−9.05**	−8.53**
60 min	−12.58**	−8.87**	−8.20**
90 min	−11.38**	−8.14**	−7.39**
120 min	−10.19**	−7.35**	−6.61**
240 min	−7.21**	−5.29**	−4.39**
480 min	−2.10**	−2.31**	−0.83
1 d	−0.21	+0.09	+0.24
2 d	−0.37	−0.04	+0.06
5 d	−0.16	−0.10	+0.03
10 d	+0.09	+0.08	+0.19
22 d	+0.18	+0.01	+0.39
Panel B: trailing window			
15 min	−5.94**	−3.47**	—
30 min	−9.11**	−5.78**	−6.14**
60 min	−11.85**	−8.41**	−8.34**
120 min	−12.52**	−8.50**	−7.84**
240 min	−10.61**	−7.69**	−6.89**
480 min	−5.68**	−4.08**	−3.39**
1 d	−0.30	+0.01	+0.31

Percentage change in MSE of $\log RV$ when one all-return control is added to HAR-CJ, on the same one-day forecasts (1,981 for Bitcoin and Ethereum, 1,864 for Litecoin); negative favors the control. Panel A: exponentially weighted realized variance with the stated half-life; Panel B: realized variance of the last stated span before the close (60 min is RV60). * and **: Holm-adjusted DM–HLN tests at 5% and 1%. —: not computed.

Bitcoin's MSE gain from EI is concentrated after late events (Table 3.17). The 145 forecasts whose last event occurred within four hours account for roughly half of the total MSE reduction. The final-hour bin contains only 35 forecasts and is too small for the reported test. Moreover, its HRMSE increases sharply, as it does in the corresponding bins of the other assets. The concentration is thus descriptive and specific to the loss; it does not imply uniform improvement

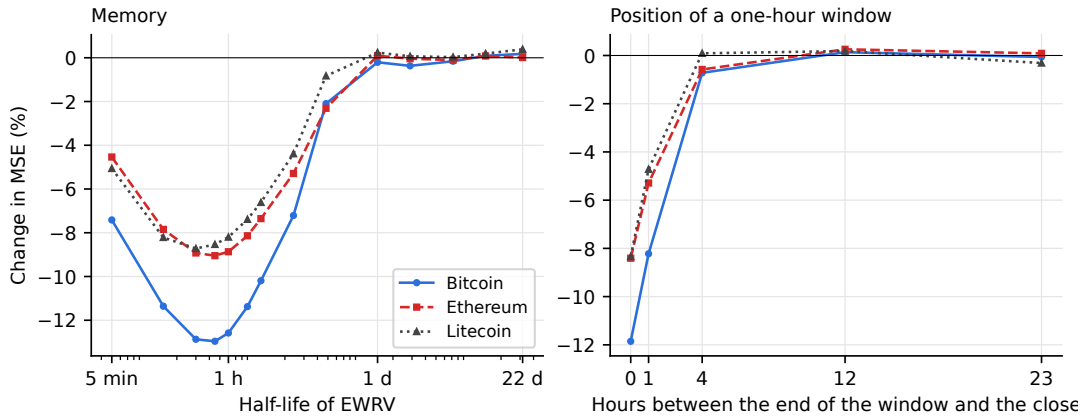


Figure 3.2: Where the gain of an all-return regressor sits. Left: EWRV enters HAR-CJ at a half-life held fixed over the whole forecast period, from five minutes to 22 days. Right: the realized variance of a one-hour window that ends the stated number of hours before the close; zero is RV60. Each point is the percentage change in MSE against HAR-CJ, main design. Windows ending two and eight hours before the close contain a day with no price change and are not shown.

after a late jump.

3.4.4 The Hour of the Jump

The timing placebo moves Bitcoin's flagged returns within their own day, preserving the daily event count and holding fitted parameters and tuned half-lives fixed. Table 3.18 reports medians over 200 draws under each relocation rule. Moving events across the whole day removes the EI gain, whereas moving them within their original hour retains most of it. This comparison tests the timing summary conditional on the fitted model.

The placebo and all-return profiles support an interpretation in terms of recent activity near the forecast origin. The placebo reports no interval around its median and does not re-estimate the process after relocation, so it cannot identify a causal effect of jump timing. Together with the control comparisons, it shows why a jump-specific interpretation requires a benchmark that already includes recent variance.

3.5 Robustness

The six sensitivities below test which parts of the main conclusion survive changes in measurement, specification and inference. Each comparison uses its

Table 3.17: The gain of the jump regressor by the time since the last jump

Asset	Time since the last event	Forecasts	Share of days (%)	Share of MSE gain (%)	Loss change (%) QLIKE	MSE	MAE	HRMSE
Bitcoin	within 1 h‡	35	1.8	16	+0.24	−13.06	−12.37	+232.78
	1 to 4 h	110	5.6	33	−6.23	−8.92	−2.40	+35.81
	4 to 24 h	324	16.4	23	−1.32	−1.75*	−1.08*	−9.14**
	1 to 3 days	567	28.6	8	+1.04	−0.36	−0.16	−5.38**
	more than 3 days	945	47.7	19	−0.03	−0.44	−0.14	−5.00**
Ethereum	within 1 h‡	38	1.9	1	+12.28	−0.20	+1.60	+156.26
	1 to 4 h	136	6.9	51	−1.56	−3.31	−1.78	+14.66
	4 to 24 h	401	20.2	26	−0.11	−0.39	−0.27	−4.55**
	1 to 3 days	628	31.7	8	+0.42	−0.06	−0.04	−2.49**
	more than 3 days	778	39.3	13	−0.00	−0.09	−0.06	−1.69**
Litecoin	within 1 h‡	30	1.6	−22	+3.35	+5.47	+11.37	+315.87
	1 to 4 h	100	5.4	48	−2.46	−2.81	−1.39	+17.00**
	4 to 24 h	330	17.7	31	−0.89	−1.10	−1.01**	−7.64**
	1 to 3 days	544	29.2	29	−0.34	−0.57	−0.30	−4.34**
	more than 3 days	860	46.1	14	−0.32	−0.13	−0.15	−2.42**

EI against HAR-CJ by the time between the last Intersection event and the forecast origin; bins fixed in advance. Share of MSE gain: the bin's part of the whole-sample reduction in squared error (it can exceed 100 or be negative). * and **: Holm-adjusted DM–HLN tests at 5% and 1%. ‡: too few forecasts for the test.

Table 3.18: Moving each jump to another time of its day: Bitcoin

New time of each flagged return	EI MSE	EI HRMSE	F MSE	F HRMSE
Events as observed	−1.16	−4.67	−0.99	−4.46
Uniform over the day	+0.04	−0.21	+0.23	+0.06
Pooled distribution over the hours	+0.02	−0.29	+0.21	+0.03
Within its own hour	−1.06	−4.43	−0.94	−4.30

Percentage change in mean loss against HAR-CJ (negative favors the regressor). First row: the primary comparison. Other rows: median over 200 draws in which every flagged return is moved within its own day and the regressors are rebuilt with the fitted parameters and the tuned half-lives held fixed.

own matched origins and benchmark. The designs were varied separately; their results do not constitute a fully crossed experiment.

3.5.1 Jump Tests, Significance Levels and Sampling Frequency

The Bitcoin gain depends on sampling and the event definition (Table 3.19). EI improves MSE with five-minute Intersection and ABD events, but not with one-minute or ten-minute Intersection events; at ten minutes its HRMSE gain remains. F is more successful under one-minute sampling. Modest changes to the Intersection tests' levels preserve Bitcoin's EI MSE gain at roughly 1% (Table 3.20). Thus level sensitivity is small within this experiment, while sampling sensitivity prevents a grid-independent claim. The equity rows are discussed in Appendix C.7.

Table 3.19: Sampling frequency and event definition: the intraday regressors against HAR-CJ

Asset	Sampling and events	Regressor	QLIKE	MSE	MAE	HRMSE
Bitcoin	1 min, Intersection	EI	−0.09	−0.29	−0.31	−0.89
		LCI	−0.11	−0.33	−0.32	−1.02
		F	−1.98	−2.79**	−1.26*	−5.07*
	5 min, Intersection	EI	−0.02	−1.12**	−0.57	−4.70**
		LCI	−0.06	−1.24**	−0.58	−5.12**
		F	+0.16	−0.88	−0.21	−4.17**
	10 min, Intersection	EI	−0.68	−0.58	−0.35	−2.17**
		LCI	−0.76	−0.69	−0.42	−2.46**
		F	−1.37	−0.99*	−0.52	−3.92**
	5 min, ABD	EI	−0.73	−1.33**	−0.86**	−3.32**
		LCI	−0.75	−1.34**	−0.87**	−3.36**
		F	−2.56*	−2.67**	−1.37**	−9.07**
S&P 500	1 min, Intersection	EI	+0.00	+0.07	+0.12	−0.07
		LCI	+0.01	+0.08	+0.14	−0.06
		F	+0.63	+0.33	+0.23	−1.31
	5 min, Intersection	EI	+0.58	+0.33	+0.10	−0.10
		LCI	+0.57	+0.31	+0.07	−0.08
		F	+0.50	+0.49	+0.35	−0.24
	5 min, ABD	EI	−0.70	−0.20	−0.01	−0.04
		LCI	−0.70	−0.20	−0.02	−0.03
		F	+0.37	+0.56	+0.33	−3.93*

Percentage change in mean loss against the HAR-CJ model of the same arm (negative favors the regressor). Each arm rebuilds the realized measures, the backbone and the events, so losses are comparable within an arm only. Process and half-life refits every 30 days on the cryptocurrencies and every 22 sessions on the S&P 500, regressions daily. * and **: significant at 5% and 1% after the Holm adjustment. The ten-minute arm of the S&P 500 is not estimable.

Table 3.20: Sensitivity to the significance levels of the jump tests

			EI				F			
Asset	Levels of the Intersection	Flags a day	QLIKE	MSE	MAE	HRMSE	QLIKE	MSE	MAE	HRMSE
Bitcoin	Main, 1% / 1% / 1%	0.46	−0.02	−1.12**	−0.57	−4.70**	+0.16	−0.88	−0.21	−4.17**
	Arm A, 2% / 1% / 2%	0.56	+0.14	−1.10**	−0.73*	−4.98**	+0.20	−1.15**	−0.64	−4.85**
	Arm B, 1% / 2% / 1%	0.48	−0.07	−1.05**	−0.54	−4.47**	+0.01	−0.98*	−0.31	−4.14**
Ethereum	Main, 1% / 1% / 1%	—	+0.26	−0.25	−0.16	−1.77**	−0.38	−0.69	−0.48	−0.31
	Arm A, 2% / 1% / 2%	0.61	−0.03	−0.28	−0.15	−1.54**	−0.29	−0.36	−0.24	+1.20
	Arm B, 1% / 2% / 1%	0.54	+0.28	−0.19	−0.11	−1.60**	−0.41	−0.69	−0.55	+0.14
Litecoin	Main, 1% / 1% / 1%	—	−0.62	−0.51	−0.23	−1.85	−0.52	−0.21	+0.04	−2.78*
	Arm A, 2% / 1% / 2%	0.52	−0.26	−0.42	−0.17	−1.33	−0.65	−0.19	+0.03	−1.54
	Arm B, 1% / 2% / 1%	0.46	−0.83	−0.61	−0.38	−1.23	−0.63	−0.33	−0.05	−2.89*

Percentage changes against HAR-CJ (negative favors the extension). * and **: significant at 5% and 1% after the Holm adjustment. The levels are those of the ABD, Lee–Mykland and sequential BNS tests. Five-minute returns, 1/7/30 cascade, Hawkes processes and tuned half-lives refitted every 30 days, daily HAR estimation. Arm A doubles the level of the ABD and sequential BNS tests and arm B that of the Lee–Mykland test. A dash marks a count of flags that the record does not store.

3.5.2 The Controls on Denser Event Sets

Denser event sets preserve the main MSE comparison but introduce HRMSE exceptions (Table 3.21). EWRV reduces Bitcoin MSE by about 13% under ABD and Union events; EI adds no significant MSE improvement. It does reduce EWRV's HRMSE by 0.57% with ABD and 1.35% with the Union. These paired-test gains qualify any claim that the controls absorb all jump information under every loss. The Union comparison of EI added to RV60 was not estimated, so its missing cells supply no negative evidence.

3.5.3 HAR Cascade and Refit Schedule

Table 3.22 compares cascades on common origins: the 1/7/30 benchmark has lower log MSE on all three cryptocurrencies, and Bitcoin's relative EI gain remains close to 1% under either cascade. This table is a cascade comparison, not a complete crossing with refit schedules. The matched daily and 30-day schedules are assessed separately in Table 3.8; their MSE sets retain EI and LCI in common, while F is retained only with daily refits. Stable EI performance therefore coexists with schedule-dependent set membership.

3.5.4 Dependence and Multiplicity

The dependence check in Table 3.23 keeps the forecasts fixed and adjusts over 16 contrasts per asset. Bitcoin's EWRV gains under MSE, MAE and HRMSE survive all reported block lengths, whereas its QLIKE gain is no longer significant

Table 3.21: The all-return controls on the sparse and the dense event sets: Bitcoin

Comparison	QLIKE	MSE	MAE	HRMSE
Intersection: 1,382 events, 1,981 forecasts				
EI against HAR-CJ	−0.06	−1.16**	−0.61	−4.67**
RV60 against HAR-CJ	−11.36**	−11.85**	−4.69**	−34.06**
EWRV against HAR-CJ	−11.53**	−12.93**	−5.39**	−37.97**
EI added to RV60	+0.20	−0.14	−0.12	−0.66
EI added to EWRV	−0.04	−0.16	−0.15	−0.36
ABD: 5,031 events, 1,981 forecasts				
EI against HAR-CJ	−0.73	−1.33**	−0.86**	−3.32**
RV60 against HAR-CJ	−11.36**	−11.85**	−4.69**	−34.06**
EWRV against HAR-CJ	−11.52**	−12.91**	−5.39**	−37.94**
EI added to RV60	−0.09	−0.07	−0.05	−0.21
EI added to EWRV	−0.11	−0.17	−0.11	−0.57*
Union: 8,397 events, 1,981 forecasts				
EI against HAR-CJ	−0.75	−1.18*	−0.77*	−2.56**
RV60 against HAR-CJ	−11.36**	−11.85**	−4.69**	−34.06**
EWRV against HAR-CJ	−11.52**	−12.91**	−5.39**	−37.94**
EI added to RV60	—	—	—	—
EI added to EWRV	−0.38	−0.09	+0.08	−1.35**

Bitcoin. Percentage change in mean loss of the first model against the second, on identical one-day forecasts (negative favors the first). EI is rebuilt on each event set; the controls do not use events. Events: intraday events of the regression sample. * and **: two-sided DM–HLN tests, Holm-adjusted, at 5% and 1%. —: not estimated on that set.

Table 3.22: The HAR cascade: 1/5/22 against 1/7/30

	1/5/22		1/7/30	
Asset	HAR-CJ	EI, %	HAR-CJ	EI, %
Bitcoin	0.5427	−1.17**	0.5216	−1.12**
Ethereum	0.4502	−0.32	0.4369	−0.25
Litecoin	0.4104	−0.51	0.4048	−0.51

Squared error of the logarithm. HAR-CJ columns give the mean loss of the benchmark; EI columns give the percentage change when EI is added. Same forecast origins and events under both cascades; process refits every 22 days. * and **: significant at 5% and 1% after the Holm adjustment.

with blocks of 100 days. Neither EI nor F adds a significant improvement to EWRV in this Intersection design. The resampling check supports the control comparison on these stored forecasts; it does not include uncertainty from repeating estimation or tuning.

Table 3.23: Serial dependence and multiplicity in the loss comparisons: Bitcoin

	Block	Holm-adjusted p-value			
Test	(days)	QLIKE	MSE	MAE	HRMSE
EWRV against HAR-CJ: MSE change -12.93%					
DM–HLN		< 0.001	< 0.001	< 0.001	< 0.001
Bootstrap	1	0.031	0.002	0.002	0.002
Bootstrap	12	0.036	0.002	0.002	0.002
Bootstrap	22	0.029	0.002	0.002	0.002
Bootstrap	100	0.060	0.002	0.002	0.003
EWRV against RV60: MSE change -1.23%					
DM–HLN		0.854	0.233	0.233	0.147
Bootstrap	1	1.000	0.816	0.816	0.493
Bootstrap	12	1.000	0.782	0.782	0.416
Bootstrap	22	1.000	0.779	0.789	0.416
Bootstrap	100	1.000	0.776	0.698	0.439
EI added to EWRV: MSE change -0.16%					
DM–HLN		0.825	0.722	0.722	0.327
Bootstrap	1	1.000	1.000	1.000	0.816
Bootstrap	12	1.000	1.000	1.000	1.000
Bootstrap	22	1.000	1.000	1.000	1.000
Bootstrap	100	1.000	1.000	1.000	1.000
F added to EWRV: MSE change +0.26%					
DM–HLN		0.928	0.928	0.565	0.303
Bootstrap	1	1.000	1.000	1.000	0.816
Bootstrap	12	1.000	1.000	1.000	1.000
Bootstrap	22	1.000	1.000	1.000	1.000
Bootstrap	100	1.000	1.000	1.000	1.000

Bitcoin, 1,981 one-day forecasts. Two-sided tests of equal mean loss. DM–HLN: Holm-adjusted over the four losses of the contrast. Bootstrap: stationary bootstrap of the loss difference with the stated block length, Holm-adjusted over the family of 16 per coin: 4 contrasts $\times$ 4 losses.

3.5.5 The Day Boundary and the HAR Specification

Table 3.24 preserves the EWRV MSE gain when Bitcoin's day ends at 06:00 or 12:00 UTC, and EI still adds no significant MSE improvement. At those boundaries, however, EI reduces EWRV's HRMSE by more than 6%. Standalone EI against HAR-CJ was not estimated at the shifted boundaries, so the table does not establish that its uncontrolled gain survives them. With semivariance terms, alone or with leverage, EI improves the underlying benchmark under MSE

and HRMSE; it supplies no significant increment after EWRV is included. The displayed cutoff and specification tests concern Bitcoin.

Table 3.24: The day boundary and the backbone: Bitcoin

Configuration	Forecasts	Benchmark MSE	EI against the backbone MSE	EI against the backbone HRMSE	EWRV against the backbone MSE	EWRV against the backbone HRMSE	EI added to EWRV MSE	EI added to EWRV HRMSE
Panel A: end of the day (UTC)								
00:00 (main)	1,980	0.5218	—	—	−12.94**	−37.98**	−0.17	−0.36
06:00	1,980	0.4977	—	—	−17.62**	−53.60**	−0.96	−6.78**
12:00	1,980	0.4856	—	—	−18.54**	−60.17**	−0.69	−6.26**
Panel B: backbone								
HAR-CJ (main)	1,981	0.5216	−1.16**	−4.67**	−12.93**	−37.97**	−0.16	−0.36
With semivariances	1,981	0.5222	−0.91**	−3.99**	−12.86**	−36.96**	−0.22	−0.41
With semivariances and leverage	1,981	0.5067	−0.87**	−3.82**	−11.26**	−32.43**	−0.19	−0.38

Bitcoin, Intersection events, daily refits. Benchmark: mean MSE of the backbone without a regressor. Percentage changes in mean loss (negative favors the first model). * and **: two-sided DM–HLN tests, Holm-adjusted, at 5% and 1%. —: not estimated.

3.5.6 Horizons of Seven and Thirty Days

No jump-arrival extension significantly improves the Bitcoin benchmark at seven or thirty days (Table 3.25). EI remains close to HAR-CJ, while F deteriorates significantly at seven days. The seven-day HRMSE gain of HAR-RV is an exception among the volatility benchmarks, not evidence for a jump-arrival regressor. The reported one-day EI improvement therefore does not establish predictive value for averages over longer horizons.

Table 3.25: Forecasts of the mean realized variance over seven and thirty days: Bitcoin

	$h = 7$				$h = 30$			
Model	QLIKE	MSE	MAE	HRMSE	QLIKE	MSE	MAE	HRMSE
HAR-CJ	0.2146	0.3290	0.4412	0.7186	0.1482	0.2880	0.4328	0.9052
HAR-RV	−0.10	−0.70	−2.02	−8.61*	−0.54	−2.56	−1.11	−8.55
DI	+0.24	+0.85	+1.45	+2.61	−1.92	−2.16	−0.10	−4.56
SI	−0.59	−0.05	+0.24	+0.07	+0.10	+0.21	+0.08	+0.80
EI	+0.66	+0.31	+0.03	−0.15	−0.17	−0.11	+0.00	−0.14
F	+6.09*	+6.84*	+4.50*	+13.72*	+25.05	+31.08	+16.27	+44.42

The target is the logarithm of the mean realized variance over the next h days. First row: mean losses of HAR-CJ; other rows: percentage change against it (negative favors the model). $n = 1,975$ at $h = 7$ and $1,952$ at $h = 30$. Five-minute Intersection events, 1/7/30 cascade, rolling 1,000-day estimation, daily refits. * and **: DM–HLN tests with $h - 1$ lags, Holm-adjusted over the four losses, at 5% and 1%. Changes above one thousand per cent are not printed.

3.6 Economic Significance Exercise

The Bitcoin forecasts are applied to the allocation and Value-at-Risk rules of Section 2.5.6 and appendix A.5. The allocation uses a 40% annual risk target, a weight capped at one, risk aversion three and a cost of ten basis points over 1,981 days. These fixed rules allow the statistical comparisons to be read alongside realized economic outcomes; they are not optimized trading policies.

The forecasts produce similar portfolio volatility, between 37.7% and 38.1%, and none improves the certainty equivalent relative to HAR-CJ (Table 3.26). F's difference is negative with an interval excluding zero; EI's interval includes zero. EWRV has the lowest point estimate of the certainty equivalent and the highest turnover, despite its statistical accuracy gains. This exercise therefore gives no evidence that the jump regressors improve the specified allocation rule.

Table 3.26: Economic significance exercise: a portfolio under a risk target, Bitcoin

Forecast	Return	Volatility	Sharpe	Drawdown	VaR	ES	CE	ΔCE	95% interval	Turnover
HAR-RV	9.54	38.06	0.25	62.62	3.04	4.57	−12.19	−0.46	[−1.42, +0.48]	25.8
HAR-CJ	9.71	37.80	0.26	61.34	3.03	4.52	−11.73	—	—	25.5
HAR-CJ-DI	9.64	37.83	0.25	61.32	3.03	4.52	−11.83	−0.09	[−0.44, +0.27]	25.6
HAR-CJ-EI	9.03	37.83	0.24	61.50	3.02	4.53	−12.44	−0.70	[−1.58, +0.08]	26.1
HAR-CJ-F	8.66	37.72	0.23	61.50	3.02	4.52	−12.68	−0.95	[−1.89, −0.06]	26.2
HAR-CJ-EWRV	7.80	37.77	0.21	62.23	3.02	4.48	−13.59	−1.86	[−4.36, +0.59]	30.8
Buy and hold	21.04	55.46	0.38	76.63	4.48	6.75	−25.10	−13.36	[−34.13, +5.28]	0.2

The rule of Section 2.5.6 with the forecast of each model: risk target 40% a year, weight capped at one, cost of 10 basis points, risk aversion 3; $n = 1,981$ days. Return, volatility and CE are annual, in per cent; Sharpe: annual return over annual volatility; drawdown: maximum, in per cent; VaR and ES: empirical, at 5%, of the daily net returns, in per cent; turnover: times wealth a year. ΔCE: CE minus that of HAR-CJ, in percentage points, with a percentile interval from a paired stationary bootstrap (mean block length 12, 10,000 draws), conditional on the forecasts. Buy and hold: the whole wealth in the asset. The cost is a scenario and not a record of executed trades.

Table 3.27 evaluates Value-at-Risk after a 250-day burn-in, leaving 1,731 targets. No model is rejected by the unconditional or conditional coverage tests at either tail level. At 1% there are only 13–15 violations, which limits the precision of this comparison. EWRV has the lowest quantile-loss ratio, while EI and F lie slightly above HAR-CJ; the displayed ratios are untested and do not establish a significant accuracy ranking. Neither economic exercise establishes an advantage for jump-based forecasts under the chosen rules.

Table 3.27: Economic significance exercise: one-day Value-at-Risk forecasts, Bitcoin

	$q = 1\%$					$q = 5\%$				
Forecast	Viol.	Rate (%)	p_{uc}	p_{cc}	Loss	Viol.	Rate (%)	p_{uc}	p_{cc}	Loss
HAR-RV	13	0.75	0.28	0.50	1.018	95	5.49	0.36	0.47	1.004
HAR-CJ	15	0.87	0.57	0.75	1.000	92	5.31	0.55	0.31	1.000
HAR-CJ-DI	13	0.75	0.28	0.50	0.999	91	5.26	0.63	0.52	1.001
HAR-CJ-EI	15	0.87	0.57	0.75	1.004	94	5.43	0.42	0.31	1.004
HAR-CJ-F	15	0.87	0.57	0.75	1.004	93	5.37	0.48	0.32	1.003
HAR-CJ-EWRV	14	0.81	0.41	0.63	0.963	85	4.91	0.86	0.91	0.989

Value-at-Risk of Definition A.4 with the forecast of each model and the empirical quantile of its past standardized returns, after a burn-in of 250 days; $n = 1,731$ days. Viol.: days on which the return falls below the Value-at-Risk; rate: their share. p_{uc} and p_{cc}: p-values of the tests of unconditional and of conditional coverage. Loss: mean quantile loss (A.20) relative to that of HAR-CJ (below one favors the model); the ratios are not tested.

Chapter 4

Conclusion

This chapter answers the research question, interprets the evidence and states the boundaries of the conclusions.

4.1 Main Results and Scope

Information about the timing and clustering of intraday jumps improves one-day Bitcoin realized-volatility forecasts relative to the bipower HAR-CJ benchmark, but the improvement is not significant once recent volatility measured from all returns enters the regression. On the main design, the closing intraday intensity lowers log MSE by 1.16% and HRMSE by 4.67%, with no significant gain under QLIKE or MAE (Table 3.7). The answer is therefore conditional on the benchmark, the loss and the sampling grid, and the five specific objectives locate its limits.

The forecasting framework fits the event process on a rolling window, recomputes the history of each regressor with the parameters of its forecast origin and compares models on common origins using information through the previous close. Within it, the coherent regressor is the conditional expected count of the next day, which for the unmarked exponential process is affine in the closing intensity, so a regression with an intercept gives the same forecast from either representation (Section 2.3.1). The two-exponential extension requires both excitation states and a conditional mark-moment assumption. The choice of summary matters empirically: the closing state improves the benchmark, whereas the daily process intensity and the integrated intraday intensity do not. The construction concerns expected event counts and does not establish an optimal predictor of realized variance.

A better fit to jump arrivals does not produce a better volatility forecast. A second exponential is preferred by the likelihood ratio at every matched refit on the three cryptocurrencies, yet none of its whole-sample loss changes against the one-exponential regressor is significant (Table 3.11). Intraday marking yields at most a 0.04% HRMSE reduction on Ethereum; the gains of the marked daily process under HRMSE are measured against the daily intensity, not against the closing intraday state.

Hawkes estimation is not necessary for the measured gain. The decayed count whose memory is tuned on forecast loss improves the same losses as the closing intensity and remains with it, and with the log-transformed count, in the daily-refit MSE confidence set (Table 3.8). Both regressors are discounted event counts that differ in how the decay is chosen (Section 2.3.2): the rolling likelihood selects a median half-life of 98 minutes and the loss criterion 130 minutes. The fitted process therefore supplies one effective rule for the memory of the count rather than information that the count lacks.

The jump intensity does not survive the all-return controls under MSE. The last-hour realized variance and the exponentially weighted realized variance reduce Bitcoin MSE by 11.85% and 12.93%, about ten times the intensity's gain, and adding the intensity to either gives a further 0.14% and 0.16% without significance under any loss (Table 3.12). Encompassing tests retain the controls over the intensity in every reported cryptocurrency panel (Table 3.15). The absorption is not complete under every criterion: the nested-model test rejects on Ethereum, and the intensity lowers the controlled HRMSE with denser event sets and shifted day boundaries (Tables 3.14, 3.21 and 3.24).

The uncontrolled gain has a narrow reach. Ethereum and Litecoin show HRMSE gains but no significant MSE gain (Table 3.9); no jump-arrival regressor improves the benchmark at seven or thirty days (Table 3.25); and the allocation and Value-at-Risk exercises establish no economic advantage under the specified rules (Tables 3.26 and 3.27).

4.2 Interpretation

The results bear on the puzzle of Section 1.1. Consistent with Andersen et al. (2007a) and Catania and Sandholdt (2019), the daily jump terms add little to the continuous components of the Bitcoin regression, while the timing of intraday jumps does improve the benchmark, as Clements and Liao (2017) report for equities. The controls change how that improvement is read: most of what the

closing intensity contributes is also carried by recent squared returns, whether or not a return is flagged as a jump.

The evidence is consistent with the intensity transmitting recent market activity near the close. The all-return information is concentrated within the last hour (Table 3.16), about half of the intensity's MSE reduction follows events in the last four hours (Table 3.17), and relocating events across the whole day removes the gain while relocation within the hour retains most of it (Table 3.18). The placebo holds the fitted parameters fixed and reports no interval around its median, so it does not identify a causal effect of jump timing.

The three questions left open in Section 1.5 receive conditional answers. The richer exponential processes that Qu et al. (2024) and Pierre (2024) motivate fit the intraday arrivals better without forecasting better on these data. A count benchmark given the same history and a tuned decay matches the Hawkes regressor, so the comparison with simple counts depends on treating both summaries alike. And a jump-specific gain requires a benchmark that already contains recent all-return variance, against which the Bitcoin MSE gain is not significant.

For a forecaster, the appropriate benchmark includes recent all-return volatility as well as daily aggregates. A gain over HAR-CJ establishes usefulness within that benchmark; additional jump-process estimation for this task is justified only by an increment over the controls. Likelihood, forecast accuracy and the value of the decision that uses the forecast are separate criteria and are evaluated separately.

4.3 Limitations and Boundaries

The process family contains one or two exponential components and power-modulated marks, and the forecasting regressions are linear in their chosen regressors. Weak intraday mark exponents and a boundary estimate limit conventional parameter inference. The general expected-count construction also requires a conditional mark-moment assumption. Results for these specifications do not settle the value of other kernels or nonlinear forecasts.

The main gain depends on measurement and evaluation. It does not persist as a significant EI MSE gain under every sampling grid or at seven- and thirty-day horizons. MCS membership changes with loss, statistic and configuration, and joint retention is not proof of equal accuracy. Likewise, insignificant controlled differences do not rule out increments too small to detect in these samples.

Observed events lie on a grid, whereas the arrival models operate in continuous

time. The simulations show finite-sample estimation bias in the daily configuration and possible goodness-of-fit rejection caused by the daily observation grid (Tables B.4 and B.7). Both intraday Bitcoin kernels also fail the distance-based adequacy tests. Their relative likelihood and forecasting performance therefore cannot validate the event model as a complete description of arrivals.

Several comparisons remain outside the evidence. The rolling Poisson test covers Bitcoin only; the richer marked ladder is evaluated at 30-day refits; standalone EI at shifted day boundaries and a fully crossed cascade-by-refit experiment were not computed. The likelihood-ratio exercise checks the marked reference case but does not estimate known-null size for Poisson against Hawkes or one exponential against two. These missing comparisons restrict the scope of the conclusions without supplying evidence for a null effect.

The cryptocurrencies are related assets observed on one exchange, and the S&P 500 sample covers a different period. Resampling conditions on stored forecasts without repeating estimation or tuning. The economic exercise uses one allocation rule and cost scenario, while tail-risk comparisons contain few extreme observations and untested loss ratios. Broader market and economic conclusions require evidence beyond these applications.

4.4 Directions for Future Research

The first line concerns the time formulation. The jumps are flagged on a five-minute grid, whereas the processes of this thesis operate in continuous time, and the simulations show that the grid alone can bias estimates and reject adequacy. Discrete-time self-exciting models are defined on the grid itself: the Hawkes process is the limit of integer-valued autoregressions (Kirchner, 2016), and the time formulation has been shown to matter for extreme events in finance (Bień-Barkowska, 2020; Bień-Barkowska and Herrera, 2026). Such models may give an equal or better regressor at lower computational cost, and they would separate discretization effects from arrival-model misspecification.

The second line extends the excitation structure. Power-law kernels decay more slowly than any exponential and represent the long memory that one or two exponentials approximate (Hardiman et al., 2013; Emmanuel Bacry et al., 2015). A multivariate process would let jumps in one asset excite jumps in another (Embrechts et al., 2011; Aït-Sahalia et al., 2015), extending the jump and cojump intensities of Clements and Liao (2017) to related cryptocurrencies. Given the result on richer exponential fits, each extension should be judged by

its forecast increment over the closing intensity and the all-return controls, not by its likelihood.

The third line relaxes the linear link between the intensity and the volatility regression. This thesis adds one regressor to a linear HAR model; nonparametric or multi-kernel specifications (K. Lee, 2024) and neural point processes (Mei and Eisner, 2017) would let the data choose the memory, and a nonlinear forecasting stage could combine the excitation states with recent all-return variance rather than add them separately. The comparison with the tuned decayed count remains necessary, since flexibility in the memory was not what separated the regressors here.

The fourth line connects jump detection with extreme value theory. A flagged jump and a peaks-over-threshold exceedance are both returns beyond a time-varying bound, and self-exciting peaks-over-threshold models combine the Hawkes intensity with a generalized Pareto distribution for the size of the exceedance (Davison and Smith, 1990; Chavez-Demoulin et al., 2005; Tomlinson et al., 2024). Applied to intraday returns, this framework would supply a distribution for the marks instead of the power modulation used here, with expected exceedance sizes and tail probabilities as candidate regressors for realized variance.

Each line should preserve matched forecast origins, direct tests against the simple exponential summary and benchmarks containing recent all-return variance. Likelihood improvements should be reported separately from forecast-loss reductions, and economic value should be tested under the decision rule in which the forecast will be used.

Appendix A

Derivations and Additional Definitions

This appendix holds three derivations that Chapter 2 uses and does not develop: the gradient of the objective of the estimation problem, the expected count of the one-exponential process, and its extension to the power-marked bi-exponential process. A further section defines the HAR-RS and LHAR-RV models, whose terms enter the benchmarks of the robustness analysis (Section 3.5). The final section defines the allocation and tail-risk measures.

A.1 Gradient of the Negative Log-Likelihood

The gradient of the objective (2.25) is given for the unmarked exponential process, $\theta = (\mu, \alpha, \beta)$. It is the gradient supplied to the solver of Appendix B.5. With $\Delta_q = t_q - t_{q-1}$, define the two states

$$\begin{aligned} R(q) &= e^{-\beta\Delta_q}\Big(1 + R(q-1)\Big), \\ D(q) &= e^{-\beta\Delta_q}\Big(\Delta_q\,(1 + R(q-1)) + D(q-1)\Big), \qquad A(1) = B(1) = 0, \end{aligned} \tag{A.1}$$

so that $\lambda^*(t_q) = \mu + \alpha\beta R(q)$ and $D(q) = \sum_{j<q}(t_q - t_j)e^{-\beta(t_q - t_j)}$. Then

$$\begin{aligned} \frac{\partial\,\mathrm{NLL}_g}{\partial\mu} &= T - \sum_{q=1}^{n}\frac{1}{\lambda^*(t_q)}, \\ \frac{\partial\,\mathrm{NLL}_g}{\partial\alpha} &= \sum_{q=1}^{n}\Big(1 - e^{-\beta(T-t_q)}\Big) - \sum_{q=1}^{n}\frac{\beta R(q)}{\lambda^*(t_q)}, \\ \frac{\partial\,\mathrm{NLL}_g}{\partial\beta} &= \alpha\sum_{q=1}^{n}(T - t_q)\,e^{-\beta(T-t_q)} - \alpha\sum_{q=1}^{n}\frac{R(q) - \beta D(q)}{\lambda^*(t_q)}. \end{aligned} \tag{A.2}$$

The three components are computed in one pass over the events, at the same cost $O(n)$ as the objective. Second derivatives are not used by the solver.

A.2 Derivation of the One-Exponential Expected Count

For the unmarked process of Section 2.3.1, the contribution of events observed by t is $k_t(s) = \mu + (\lambda_t - \mu)e^{-\beta(s-t)}$. Conditional expectation of the event increment inside (t, s) equals $m_t(u)\,\mathrm{d}u$ by the compensator identity, and the unmarked kernel integrates to $\eta = \alpha$. Hence

$$m_t(s) = k_t(s) + \int_t^s \eta\beta e^{-\beta(s-u)} m_t(u)\,\mathrm{d}u. \tag{A.3}$$

Differentiating and using $k_t'(s) = -\beta(k_t(s) - \mu)$ gives

$$m_t'(s) = -\beta(1-\eta)m_t(s) + \beta\mu, \qquad m_t(t) = \lambda_t. \tag{A.4}$$

Under $\eta < 1$, the fixed point is $\lambda_\infty = \mu/(1-\eta)$, so solving this equation gives (2.3.1). Integration over $(t, t+h]$ gives (2.28). The same solution and count hold over a finite horizon for every $\eta \neq 1$, with λ_∞ then only the constant $\mu/(1-\eta)$; it is the stationary mean intensity only when $\eta < 1$. At $\eta = 1$ the equation gives $m_t(t+u) = \lambda_t + \beta\mu u$ and the count $\lambda_t h + \beta\mu h^2/2$ (Appendix A.3).

A.3 Expected Count of the Power-Marked Bi-Exponential Process

The result extends Section 2.3.1 to Equation (2.16). The state now has two components. Observed marks enter the initial state individually; prediction of future excitation also requires a model for future marks.

Specification and assumptions

Let parameters be fixed over the prediction horizon and let $\xi_q > 0$. Assume the conditional mean of ξ^δ at an arrival, given the history just before that arrival, is a constant $\nu_\delta < \infty$. Independent, identically distributed marks independent of the preceding history are sufficient. In random-measure notation, this assumption

means

$$\mathbb{E}\Big[\int \xi^{\delta} N(\mathrm{d}s, \mathrm{d}\xi) \,\Big|\, \mathcal{H}_{s^-}\Big] = \nu_\delta \lambda^*(s)\,\mathrm{d}s.$$

Assume finite first moments on the forecast window. Define $w_k = \alpha_k \nu_\delta$ and $\eta = w_1 + w_2$. For the stationary formula below, assume $\eta < 1$. A sample mean of marks is a plug-in estimate of ν_δ only under a suitable mark model; persistent financial marks do not automatically satisfy the conditional-moment assumption.

Write $\lambda^*(s) = \mu + x_1(s) + x_2(s)$, where

$$x_k(s) = \alpha_k \beta_k \sum_{t_q < s} \xi_q^{\delta} e^{-\beta_k (s - t_q)}, \qquad k = 1, 2. \tag{A.5}$$

At the forecast origin use the post-event vector $x_t = (x_1(t^+), x_2(t^+))^\top$, so an event at t is already observed. The state dynamics are

$$\mathrm{d}x_k(s) = -\beta_k x_k(s)\,\mathrm{d}s + \alpha_k \beta_k \int \xi^{\delta} N(\mathrm{d}s, \mathrm{d}\xi). \tag{A.6}$$

Conditional mean and integrated count

Proposition A.1 (Expected count with power marks and two decay rates)**.** *Put* $a = (w_1\beta_1, w_2\beta_2)^\top$, $\mathbf{1} = (1,1)^\top$, *and*

$$A = -\operatorname{diag}(\beta_1, \beta_2) + a\mathbf{1}^\top, \qquad \lambda_\infty = \frac{\mu}{1-\eta}, \qquad \bar{x}_\infty = \lambda_\infty (w_1, w_2)^\top.$$

Under the stated assumptions,

$$\mathbb{E}\Lambda_t(h) = \lambda_\infty h + v^\top (x_t - \bar{x}_\infty), \qquad v^\top = \mathbf{1}^\top A^{-1}(e^{Ah} - I). \tag{A.7}$$

Proof. Let $\bar{x}(u) = \mathbb{E}[x(t+u) \mid \mathcal{H}_t]$ for $u > 0$, and $\bar{x}(0) \coloneqq \bar{x}(0^+) = x_t$, since the states are left-continuous and x_t is the post-event state. Taking conditional expectations in Equation (A.6), using the conditional mark moment and the tower property, gives

$$\bar{x}'(u) = A\bar{x}(u) + \mu a, \qquad \bar{x}(0) = x_t. \tag{A.8}$$

The determinant is $\beta_1\beta_2(1-\eta) > 0$. Substitution shows $A\bar{x}_\infty + \mu a = 0$, so $\bar{x}(u) = \bar{x}_\infty + e^{Au}(x_t - \bar{x}_\infty)$. The compensator identity of Section 2.3.1 gives

$$\mathbb{E}\Lambda_t(h) = \int_0^h \{\mu + \mathbf{1}^\top \bar{x}(u)\}\,\mathrm{d}u.$$

Since $\mu + \mathbf{1}^\top \bar{x}_\infty = \lambda_\infty$ and $\int_0^h e^{Au}\,\mathrm{d}u = A^{-1}(e^{Ah} - I)$, integration gives Equation (A.7). □

The count is affine in the vector of closing states. It is generally not affine in their sum, the closing intensity $\lambda_t = \mu + \mathbf{1}^\top x_t$. Indeed,

$$\begin{aligned} v^\top &= h\mathbf{1}^\top + \frac{h^2}{2}\mathbf{1}^\top A + O(h^3), \\ \mathbf{1}^\top A &= (w_1\beta_1 + w_2\beta_2)\mathbf{1}^\top - (\beta_1, \beta_2). \end{aligned} \tag{A.9}$$

When both states can vary and $\beta_1 \neq \beta_2$, their weights differ at second order. Two histories can have the same closing intensity and different conditional expected counts. The scalar reduction is recovered when only one component is active, or when the decays coincide.

Finite-horizon form without a stationarity denominator

An augmented linear system avoids division by $1 - \eta$ and the inverse of A, which is singular at $\eta = 1$. Define $c(0) = 0$ and $c'(u) = \mu + \mathbf{1}^\top \bar{x}(u)$. Then

$$\begin{pmatrix} \bar{x}(h) \\ 1 \\ c(h) \end{pmatrix} = \exp\left[h \begin{pmatrix} A & \mu a & 0_{2\times 1} \\ 0_{1\times 2} & 0 & 0 \\ \mathbf{1}^\top & \mu & 0 \end{pmatrix} \right] \begin{pmatrix} x_t \\ 1 \\ 0 \end{pmatrix}. \tag{A.10}$$

Under the finite-moment and mark assumptions, $c(h) = \mathbb{E}\Lambda_t(h)$ for a finite horizon and every value of η, including $\eta = 1$. This mathematical representation does not assert that the empirical implementation estimates a critical model. For one component with $\eta = 1$, the mean intensity is $\lambda_t + \beta\mu u$ and the count is $\lambda_t h + \beta\mu h^2/2$.

Setting $\delta = 0$ gives $\nu_\delta = 1$ and the unmarked bi-exponential result. Setting $\alpha_2 = 0$ gives the marked exponential result with branching ratio $\alpha_1\nu_\delta$; additionally setting $\delta = 0$ gives Section 2.3.1. The fixed-parameter condition matters for forecasting regressions: affine equivalence need not survive parameter changes across rows, and taking a logarithm is not an affine transformation.

The unmarked regressors used in the empirical comparison

For the unmarked case let $R_{k,t} = \sum_{t_q \le t} e^{-\beta_k(t - t_q)}$, $d_0 = \lambda_\infty h - v^\top \bar{x}_\infty$, and $d_k = v_k\alpha_k\beta_k$. Then $\mathbb{E}\Lambda_t(h) = d_0 + d_1 R_{1,t} + d_2 R_{2,t}$. The empirical names below

refer to this unmarked case; the general formula does not change what the recorded experiments estimated.

Definition A.1 (Regressors of the two-exponential process)**.** With component 1 the fast component and $h = 1$,

$$EIbi_t = d_0 + d_1 R_{1,t} + d_2 R_{2,t}, \tag{A.11}$$

$$EIbi_t^+ = \alpha_1 \beta_1 R_{1,t}, \tag{A.12}$$

$$EIbi_t^- = \alpha_2 \beta_2 R_{2,t}. \tag{A.13}$$

The model 2EIbi enters $EIbi_t^+$ and $EIbi_t^-$ separately.

EIbi combines the states using weights implied by the fitted event model. The two-regressor version lets the forecasting regression choose their weights, while the separate components assess each time scale. With fixed parameters and an intercept, the two-regressor specification contains the count regressor in its design space.

A.4 The HAR-RS and LHAR-RV Models

Definition A.2 (Semivariance and leverage extensions)**.** Let RS_t^+ and RS_t^- be the realized semivariances of day t, the sums of the squared positive and of the squared negative intraday returns, so that $RV_t = RS_t^+ + RS_t^-$ (Barndorff-Nielsen et al., 2008), and let $r_t^- = \min\{r_t, 0\}$ with r_t the return of day t. The HAR-RS model replaces the daily term of (2.32) by the two semivariances (Patton and Sheppard, 2015),

$$\begin{aligned} \log RV_t = \phi_0 + \phi^+ \log RS_{t-1}^+ + \phi^- \log RS_{t-1}^- \\ + \phi_w \log RV_{t-1}^{(w)} + \phi_m \log RV_{t-1}^{(m)} + \varepsilon_t, \end{aligned} \tag{A.14}$$

and the LHAR-RV model adds to (2.32) the cascade of negative returns (Corsi and Renò, 2012),

$$\begin{aligned} \log RV_t = \phi_0 + \phi_d \log RV_{t-1} + \phi_w \log RV_{t-1}^{(w)} + \phi_m \log RV_{t-1}^{(m)} \\ + \ell_d \, r_{t-1}^- + \ell_w \, r_{t-1}^{-(w)} + \ell_m \, r_{t-1}^{-(m)} + \varepsilon_t. \end{aligned} \tag{A.15}$$

In (A.14) the variation carried by negative and by positive returns may have different effects on the next day; in (A.15) it is the size of past negative returns that enters, at the three horizons of the cascade. The two models use no

point process. Their terms are added to HAR-CJ in the robustness analysis of Section 3.5.5.

A.5 Economic Definitions

Let $\widehat{V}_t$ be a variance forecast in decimal units, R_t the simple return and σ^* the annual risk target. A forecast in squared percentage points is divided by 100^2 before it enters the allocation. The weight and net return are

$$w_t = \min\{1, \sigma^*/\sqrt{365\widehat{V}_t}\}, \tag{A.16}$$

$$R_t^p = w_t R_t - c|w_t - w_{t-1}|, \qquad w_0 = 0. \tag{A.17}$$

The balance is held in zero-yield cash. Trading takes place at the forecasting close, and intraday weight drift is ignored. The cost c is a scenario, not a record of execution.

For risk aversion $\gamma > 0$, the annual certainty equivalent is

Definition A.3 (Certainty-equivalent return)**.**

$$CE = 365\left(\bar{R}^p - \frac{\gamma}{2}s_R^2\right). \tag{A.18}$$

It evaluates the mean-variance tradeoff of the realized net returns. Differences between forecasts use a paired stationary-bootstrap interval, conditional on those forecasts. Annual return, volatility, their ratio at zero cash yield, and maximum drawdown summarize the same portfolio; the reported empirical tail measures describe its net returns.

Value-at-Risk is the lower return quantile, with negative values denoting adverse returns:

Definition A.4 (Value-at-Risk)**.**

$$\mathbb{P}(r_t \leq VaR_{q,t} \mid \mathcal{F}_{t-1}) = q. \tag{A.19}$$

Under $r_t = \sqrt{V_t}Z_t$ with a stable standardized-return distribution, the forecast is $\widehat{VaR}_{q,t} = \sqrt{\widehat{V}_t}\,\hat{z}_{q,t}$, where $\hat{z}_{q,t}$ is the empirical quantile of earlier standardized returns from the same forecasting model. Only past observations enter that estimate, and the evaluation excludes the initial 250-day burn-in.

For a forecast quantile v_t, the quantile loss is

$$\rho_q(r_t, v_t) = (q - \mathbf{1}\{r_t < v_t\})(r_t - v_t). \tag{A.20}$$

Its sample mean evaluates quantile accuracy. The coverage tests of Section 2.5.6 instead examine violation frequency and clustering. The ratios reported in Table 3.27 are descriptive; a favorable ratio is not a tested difference in quantile accuracy.

Appendix B

Simulation and Numerical Validation

This appendix tests the methods of Chapter 2 on simulated data. Each section states a theoretical result for a process with known parameters and reports an exercise in which the methods, as implemented, are compared with it: the mean and the variance of the count of a stationary process, the unit exponential law of the durations under the time-rescaling theorem, the recovery of the true parameters by the estimator, the test statistics at estimated parameters, the expected count and the likelihood ratio. The exercises are evidence that the tools work on the cases of Appendix B.1. They say nothing about whether the jump times of Chapter 3 follow a Hawkes process. The appendix also states three methods that the chapters use on the data: the simulator (Appendix B.2), the numerical method of the estimation (Appendix B.5) and the test of fit (Appendices B.4 and B.6).

B.1 Design

The checks proceed from simulation and likelihood identities to time rescaling, estimation, fitted-model diagnostics and expected counts. This order prevents a later tool from serving as the sole reference for an earlier one.

All checks use the unmarked exponential process. For each of thirteen configurations, 1,000 independent paths are simulated after burn-in on $[0, 10{,}000]$. Moment checks use the full bank. The remaining checks use the first 100 paths of four configurations at $T = 1{,}000$ and $T = 10{,}000$. Bootstrap replicates and conditional-count paths are simulated separately because their initial states or

parameters depend on the check.

Every configuration has stationary intensity one, $\mu = 1 - \eta$, so that a window of length T holds T events in expectation. Nine configurations form the grid $\eta \in \{0.1, 0.5, 0.9\} \times \beta \in \{0.1, 1, 10\}$. Two have the parameters of the processes fitted to the full sample of Bitcoin (Table 3.4): the intraday process of the Intersection events and the daily process. They are called the intraday and the daily configuration. One further configuration, $(\eta, \beta) = (0.299, 14.9)$, lies between the grid and the intraday fit and is used in step 1 only. The thirteenth is a homogeneous Poisson process. The four configurations of steps 2 to 5 are the grid cells $(\eta, \beta) = (0.5, 1)$ and $(0.9, 1)$, the intraday and the daily one. The normalization loses nothing, by the following property.

Proposition B.1 (Change of time unit)**.** *Let N be the process of (2.17) with parameters (μ, α, β) and no marks, and let $c > 0$. The process $\tilde{N}$ with event times $\tilde{t}_q = c\, t_q$ is of the same family, with parameters $(\mu/c,\ \alpha,\ \beta/c)$.*

Proof. An interval of length Δ at $\tilde{t} = ct$ is the image of an interval of length Δ/c at t, so $\tilde{\lambda}^*(\tilde{t}) = \lambda^*(\tilde{t}/c)/c$. Substituting (2.17), $\tilde{\lambda}^*(\tilde{t}) = \mu/c + \sum_{\tilde{t}_q < \tilde{t}} \alpha\, (\beta/c)\, e^{-(\beta/c)(\tilde{t} - \tilde{t}_q)}$. □

The branching ratio is unchanged, and so is every statistic of this appendix that is a function of the rescaled durations or of a likelihood ratio. The Bitcoin fits are therefore expressed in units of their mean spacing between events, about 2.1 days for the intraday events and 3.0 days for the jump days, which gives $(\eta, \beta) = (0.176, 65.0)$ for the intraday configuration and (0.720, 0.119) for the daily one. In the intraday units a five-minute interval is 0.0016; in the daily units a day is 0.334.

Two conventions apply to every table. A rate estimated from R independent paths at a nominal value a has standard error $\sqrt{a(1-a)/R}$ and is accepted when it lies within three standard errors of a. With the 100 paths of steps 2 to 5 the standard error of a rate near 5% is 0.022, so the tables can show that a tool is not grossly wrong and cannot certify a size. And each check is accompanied, where its cost is small, by a negative control: a deliberate error that the check must detect. A check that cannot fail is not evidence. Because the steps share their paths, they are not independent confirmations of one another, and results at two window lengths are correlated.

The lowest branching ratio among the four configurations of steps 2 to 5 is 0.18, that of the intraday fit, and the marked and two-exponential processes are not examined.

B.2 The Simulator

The paths of every step are drawn by the thinning algorithm, which is also the algorithm behind the bootstrap references of Chapter 2 and of Chapter 3.

Proposition B.2 (Thinning)**.** *The thinning construction is due to Yosihiko Ogata (1981). Let $M \geq \lambda^*(s)$ for all s in an interval starting at the current time, and generate a candidate after an* $\mathrm{Exp}(M)$ *waiting time. Accepting the candidate at t' with probability $\lambda^*(t')/M$, and discarding it otherwise, yields points whose conditional intensity is λ^*. The construction requires only that such an M be available, not that λ^* have any particular form.*

For every kernel used here the intensity is non-increasing between events, so the bound may be taken as the post-jump value $\lambda^*(s^+)$ at the last event or rejected candidate, the tightest bound available without further computation, and the one Algorithm 2 uses. A candidate that is rejected contributes no impact, which is the convention $\omega = 0$ for thinned points.

Algorithm 2 Ogata thinning for a process with a decaying conditional intensity

Require: baseline μ; kernel with $K(0) = 1$; the impact ω an accepted event carries; horizon T
Ensure: event times $\mathcal{T} = \{t_q\}$ on $(0, T]$ with conditional intensity λ^*

$s \leftarrow 0, \quad S \leftarrow 0, \quad \omega^{\text{last}} \leftarrow 0, \quad \mathcal{T} \leftarrow \varnothing$
while $s < T$ **do**
  $M \leftarrow \mu + S + \omega^{\text{last}}$ ▷ the post-jump bound $\lambda^*(s^+)$
  draw $U_1 \sim \mathrm{Unif}(0,1)$; $\quad t' \leftarrow s - \log(U_1)/M$
  if $t' > T$ **then**
    break
  end if
  $S \leftarrow K(t'-s)\big(S + \omega^{\text{last}}\big)$ ▷ decay the state onto t', as in (2.2.4)
  draw $U_2 \sim \mathrm{Unif}(0,1)$
  if $U_2 \leq (\mu + S)/M$ **then**
    $\mathcal{T} \leftarrow \mathcal{T} \cup \{t'\}; \quad \omega^{\text{last}} \leftarrow \omega$ ▷ accepted: its impact enters the next bound
  else
    $\omega^{\text{last}} \leftarrow 0$ ▷ rejected: a thinned point carries no impact
  end if
  $s \leftarrow t'$
end while
return $\mathcal{T}$

Remark (Burn-in). A draw started from an empty history opens below its stationary regime, so a replicate meant to stand in for an observed window

discards a burn-in before the retained span begins. A bootstrap built without it is short of events by construction, and the shortfall pushes every statistic in the same direction.

B.3 The Count of a Stationary Process

The first reference is the law of the number of events in a window, $N_T \coloneqq N((0,T])$, which is known in closed form for the stationary exponential process.

Proposition B.3 (Moments of the count)**.** *Under Section 2.3.1, the stationary unmarked process of* (2.17) *satisfies*

$$\mathbb{E}[N_T] = \lambda_\infty T. \tag{B.1}$$

Its covariance density, $c(\tau)\,\mathrm{d}t\,\mathrm{d}\tau = \mathrm{Cov}\left(N(\mathrm{d}t), N(\mathrm{d}t+\tau)\right)$ *for* $\tau > 0$*, is*

$$c(\tau) = \lambda_\infty \frac{\eta\beta\,(2-\eta)}{2\,(1-\eta)}\, e^{-\beta_* \tau}, \qquad \beta_* = \beta(1-\eta), \tag{B.2}$$

and the variance of the count is

$$\mathrm{Var}(N_T) = \frac{\lambda_\infty}{(1-\eta)^2}\left[T - \eta(2-\eta)\,\frac{1-e^{-\beta_* T}}{\beta_*}\right]. \tag{B.3}$$

The mean follows from taking expectations in (2.17) under stationarity. The covariance density is the result of Hawkes (1971) for the exponential kernel, and the variance follows from it by $\mathrm{Var}(N_T) = \lambda_\infty T + 2\int_0^T (T-u)\,c(u)\,\mathrm{d}u$, which holds for a simple stationary process.

Two consequences organize the check. First, $\mathrm{Var}(N_T)/T \to \lambda_\infty/(1-\eta)^2$, and the standardized count $(N_T - \lambda_\infty T)/\sqrt{T}$ is asymptotically normal with that variance (E. Bacry et al., 2013). Second, the mean (B.1) depends on μ and η alone. A Poisson process of rate λ_∞ has the same mean, and a simulator with a wrong decay would pass a check of the mean. The variance does depend on β, through the finite-window term of (B.3), and it exceeds the Poisson variance $\lambda_\infty T$ by the factor $(1-\eta)^{-2}$ in the limit. It is therefore the variance that ties the three parameters of the simulator to their meaning, and the check requires both moments.

Table B.1 reports the comparison. Of the 96 differences of the twelve Hawkes configurations, none exceeds three standard errors and four exceed two, the

Table B.1: Mean, variance and normality of the simulated count against their theoretical values

		Mean count, by T				Variance of the count, by T				Normality
η	β	10	100	1,000	10,000	10	100	1,000	10,000	p
0.1	0.1	+0.02	+0.94	+1.27	+1.38	−1.21	+0.35	−1.06	−0.05	0.118
0.1	1	−2.29	−0.19	−0.93	−0.09	−0.25	+1.09	+0.53	+0.06	0.886
0.1	10	+0.49	+1.67	−0.47	−1.55	−0.59	−0.66	−1.73	+0.01	0.022
0.5	0.1	−1.38	−1.80	+0.51	+0.16	+0.03	−1.55	−1.40	+0.06	0.387
0.5	1	+0.04	−0.22	−0.23	+0.73	+0.34	+0.06	−0.64	+0.63	0.295
0.5	10	−1.65	+0.32	+0.87	+0.11	−0.97	−0.03	+0.25	−0.48	0.655
0.9	0.1	+1.38	+0.65	+0.92	−0.17	+2.10	+0.63	+0.57	+0.05	0.188
0.9	1	−1.12	−1.47	−0.88	+0.01	−0.98	−1.85	+0.29	+2.23	0.118
0.9	10	−1.73	−1.04	−0.64	+1.86	−1.93	−1.86	+1.12	−0.57	0.483
0.299	14.9	−0.73	−0.78	+0.22	+0.76	+0.59	+1.04	+1.09	+0.03	0.237
0.720	0.119	+0.18	+0.81	+1.32	+2.07	+0.69	+0.97	+1.02	−1.75	0.064
0.176	65.0	−1.80	−0.26	−0.29	−1.40	−1.01	+1.73	−1.28	+1.08	0.059
0	—	−1.44	−0.98	+0.22	−1.11	−0.50	+0.35	−0.33	−0.11	0.164

Unmarked exponential Hawkes process with stationary intensity one; time is measured in mean spacings between events. Rows: the nine cells of the grid; two configurations at the scale of the application, of which the second is the full-sample daily fit on Bitcoin; the full-sample fit on Bitcoin of the intraday process of the Intersection events; a Poisson process. 1,000 paths per configuration, simulated on $[0, 10{,}000]$ after a burn-in and restricted to $[0, T]$. Entries: the simulated moment minus its theoretical value, in standard errors. Last column: Kolmogorov–Smirnov p-value of the standardized count at $T = 10{,}000$ against the normal law. Last row: a homogeneous Poisson process, against the Poisson moments. Negative controls: without the burn-in the mean count on $[0, 100]$ at $\eta = 0.9$, $\beta = 0.1$ lies 51 standard errors below its value; the Poisson paths, judged against the variance of the configuration $\eta = 0.5$, $\beta = 1$, lie 67 to 70 standard errors below it, with the mean within 1.1.

largest being 2.29, and normality of the standardized count is not rejected at the 1% level in any configuration. Both negative controls are detected. Without the burn-in the simulated count falls far below its mean. The Poisson paths pass the check of the mean and fail that of the variance, which is the point made above: the mean alone does not identify the decay.

Identities of the Likelihood

The moments of the count exercise the simulator and not the likelihood. The likelihood is checked by four identities that hold path by path and need no reference distribution. The recursion of Section 2.2.4 must return the same log-likelihood as the direct double sum over pairs of events that it replaces. The closed-form compensator in (2.25) must equal the numerical integral of the intensity over $[0, T]$. The gradient of Appendix A.1 must equal the finite differences of the objective, and the numerical Hessian used for the standard errors must equal the finite differences of that gradient. And at $\alpha = 0$ the objective must reduce to the Poisson log-likelihood $n \log \mu - \mu T$, which at $\mu = n/T$ is $n \log(n/T) - n$, the value used in Section 2.2.6. The direct sum and the numerical integral are computed by code that shares nothing with the implementation under test.

Table B.2: Identities of the likelihood: largest relative deviation across simulated paths

Identity	Tolerance	True parameters	Perturbed parameters
Recursion against the direct double sum	10^{-10}	1.6×10^{-15}	4.0×10^{-15}
Closed-form compensator against quadrature	10^{-8}	2.9×10^{-14}	4.2×10^{-14}
Analytic gradient against finite differences	10^{-5}	2.4×10^{-7}	3.5×10^{-8}
Standard errors against the Hessian of the analytic gradient	10^{-3}	5.7×10^{-5}	1.6×10^{-5}
Poisson limit at $\alpha = 0$	10^{-12}	4.1×10^{-16}	—

20 paths of 1,000 events in each of the first eleven configurations of Table B.1. Perturbed parameters: $(1.3\mu,\ 0.8\alpha,\ 1.5\beta)$. The fourth identity is evaluated on the 359 of 440 path-points at which the Hessian is negative definite; at the others, none of which is the maximizer of its path, no standard error exists.

The deviations of Table B.2 are at the level of rounding for the recursion, the compensator and the Poisson limit, and at the level of the finite-difference error for the gradient and the standard errors. One fact matters for the later steps. At a point that is not the maximizer of its path, the Hessian of the log-likelihood of a path of 1,000 events is often not negative definite, most often with a slow decay or a low branching ratio. The information matrix of Appendix B.5 is therefore computed as a mean over paths, and standard errors are computed at

the estimate, where no such failure occurred.

B.4 Time Rescaling at the True Parameters

Proposition B.4 (Time rescaling)**.** *By the time-change theorem as stated by Daley and Vere-Jones (2003), if N has compensator Λ^*, the transformed times $\tilde{t}_q = \Lambda^*((0, t_q])$ form a homogeneous Poisson process of unit rate in the sense of Section 2.2.1, so the interarrival times*

$$\tau_q = \Lambda^*\Big((t_{q-1}, t_q]\Big), \qquad t_0 := 0, \tag{B.4}$$

are i.i.d. unit exponential.

Proposition B.4 turns goodness of fit into a one-dimensional distributional problem: a fitted process is adequate to the extent that its rescaled durations (B.4) are i.i.d. Exp(1). The *marginal* law is read by two distance statistics, Kolmogorov–Smirnov and Cramér–von Mises, and the *independence* by a Ljung–Box statistic on the durations. They are reported side by side and not combined into one verdict.

By Proposition B.4 the durations (B.4), computed with the compensator at the true parameters, are i.i.d. unit exponential, exactly and for every sample size. Let $u_q = 1 - e^{-\tau_q}$, uniform on $(0, 1)$ under that law, with order statistics $u_{(1)} \le \cdots \le u_{(n)}$ and empirical distribution function F_n. The two distance statistics are

$$\begin{aligned} D_n &= \sup_{0 \le u \le 1} \Big| F_n(u) - u \Big|, \\ W^2 &= n \int_0^1 \Big(F_n(u) - u \Big)^2 \, \mathrm{d}u = \frac{1}{12n} + \sum_{i=1}^{n} \left(u_{(i)} - \frac{2i-1}{2n} \right)^2 . \end{aligned} \tag{B.5}$$

When the hypothesis is fully specified their limit laws do not depend on it: $\mathbb{P}(\sqrt{n}\, D_n \le x) \to 1 - 2\sum_{k \ge 1} (-1)^{k-1} e^{-2k^2 x^2}$ (Massey, 1951), and W^2 converges to the law tabulated by Anderson and Darling (1952). The 5% critical values are 1.358 and 0.461. Independence is examined with the statistic of Ljung and Box (1978) on the first m autocorrelations $\hat{\rho}_k$ of the durations, $Q = n(n+2) \sum_{k=1}^{m} \hat{\rho}_k^2 / (n-k)$, referred to χ^2_m, with $m = 20$.

The check is that each test rejects at its nominal rate, at the 1%, 5% and 10% levels. It uses no estimate: the transform is the one under test, and the tests are those of standard numerical libraries.

Table B.3: Rejection rates of the exponential and independence tests on durations rescaled at the true parameters

				Kolmogorov–Smirnov			Cramér–von Mises			Ljung–Box		
Configuration	η	β	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
Grid	0.5	1	1,000	0.00	0.02	0.08	0.01	0.02	0.05	0.03	0.10	0.16
			10,000	0.01	0.06	0.10	0.03	0.05	0.08	0.00	0.06	0.11
Grid	0.9	1	1,000	0.02	0.05	0.10	0.02	0.05	0.11	0.02	0.04	0.08
			10,000	0.00	0.06	0.15	0.00	0.06	0.12	0.01	0.08	0.14
Intraday	0.176	65.0	1,000	0.03	0.06	0.07	0.00	0.06	0.07	0.01	0.05	0.12
			10,000	0.01	0.08	0.11	0.01	0.06	0.12	0.01	0.05	0.10
Daily	0.720	0.119	1,000	0.00	0.04	0.10	0.00	0.02	0.11	0.00	0.02	0.08
			10,000	0.01	0.04	0.06	0.00	0.04	0.07	0.00	0.05	0.11

Unmarked exponential Hawkes process with stationary intensity one; time is measured in mean spacings between events. Intraday and Daily: the branching ratio and decay of the full-sample fits on Bitcoin of the intraday process of the Intersection events and of the daily process. Share of 100 paths on which each test rejects at the level of the column; tests of standard numerical libraries, Ljung–Box with 20 lags. With 100 paths the standard error of a rate near 5% is 0.022. Negative control: with the compensator multiplied by 1.05 the Kolmogorov–Smirnov test rejects at 5% on 98% to 99% of the paths at $T = 10{,}000$.

Table B.3 reports the rejection rates. At the 5% level they lie between 0.02 and 0.10 over the 24 cells, none beyond three standard errors of the nominal value, and the negative control is rejected on nearly every path.

B.5 The Estimator

The reference for the estimator is the limit (2.2.4). Its covariance is the inverse of the information per unit time, which has the following form for a point process.

Proposition B.5 (Information matrix)**.** *For a stationary process with intensity $\lambda^*(t;\theta)$, twice differentiable in θ, the score at the true value has mean zero and*

$$\begin{aligned} \mathcal{I}(\theta^\circ) &= \mathbb{E}\left[\frac{1}{\lambda^*(t)}\,\frac{\partial\lambda^*(t)}{\partial\theta}\,\frac{\partial\lambda^*(t)}{\partial\theta^\top}\right] \\ &= \lim_{T\to\infty}\frac{1}{T}\,\mathbb{E}\left[\frac{\partial^2 \mathrm{NLL}_g}{\partial\theta\,\partial\theta^\top}\right], \end{aligned} \tag{B.6}$$

all derivatives being taken at θ°.

The two expressions and the regularity conditions under which the limit (2.2.4) holds are those of Yoshiko Ogata (1978, Theorem 5); see also Daley and Vere-Jones (2003).

For the unmarked exponential process, with the sums $R(t) = \sum_{t_q<t} e^{-\beta(t-t_q)}$

and $D(t) = \sum_{t_q<t}(t - t_q)e^{-\beta(t-t_q)}$ of Appendix A.1,

$$\frac{\partial\lambda^*}{\partial\mu} = 1, \qquad \frac{\partial\lambda^*}{\partial\alpha} = \beta R(t), \qquad \frac{\partial\lambda^*}{\partial\beta} = \alpha\Big(R(t) - \beta D(t)\Big). \tag{B.7}$$

The derivative with respect to the decay is proportional to α, so the entries of $\mathcal{I}$ in the row and column of β vanish as $\alpha \to 0$: the information on the decay is of order η^2, and at the Poisson point the matrix is singular. This is the statement behind two facts used in the thesis, that the decay is poorly determined when self-excitation is weak, and that the test of $\alpha = 0$ has no standard limit (Section 2.2.6).

The Numerical Method

Problem (2.2.4) is solved by L-BFGS-B with box constraints (Byrd et al., 1995; Zhu et al., 1997). The objective and gradient use linear-time recursions. Because the likelihood is not jointly concave in the decay, five initial half-lives are spaced logarithmically between one sampling interval and the estimation-window length. Each starts at branching ratio 0.5 with a baseline matching the observed rate; the fit with the highest likelihood is retained. This is a numerical search strategy, not a guarantee of a global maximum.

On simulated paths, no selected estimate has a likelihood below that at the true parameters. The start plan matches a search initialized at the truth on 398 of 400 paths at 1,000 events and on 91 of 100 intraday paths at 10,000 events; when they differ, the plan attains the higher likelihood. On twelve Bitcoin fitting cases it matches an expanded deterministic start plan, while five random starts reach a poorer Poisson solution in four cases.

Recovery of the Parameters

Recovery is evaluated by relative bias, the empirical standard deviation relative to the inverse-information prediction, and coverage of nominal 95% observed-information intervals (Table B.4). The two sample lengths distinguish finite-sample behavior from the asymptotic approximation.

At 10,000 expected events, the empirical standard deviations are between 0.83 and 1.07 times the asymptotic values, and coverage ranges from 91 to 99 of 100 paths. The daily configuration retains some bias. At 1,000 expected events its baseline is overestimated by 32%, its branching ratio underestimated by 12%, and its decay overestimated by 19%; coverage of the first two parameters is only

Table B.4: Recovery of the parameters by maximum likelihood on simulated paths

				$\hat{\mu}$			$\hat{\alpha}$			$\hat{\beta}$		
Configuration	η	β	T	Bias	Ratio	Cov.	Bias	Ratio	Cov.	Bias	Ratio	Cov.
Grid	0.5	1	1,000	+0.5	0.96	97	−1.0	0.97	96	+0.0	0.87	96
			10,000	−0.4	0.83	98	+0.5	0.90	96	−0.4	0.86	99
Grid	0.9	1	1,000	+2.3	1.03	93	−0.9	1.13	92	+0.4	1.07	95
			10,000	+0.0	1.00	95	−0.2	1.03	94	−0.3	1.07	94
Intraday	0.176	65.0	1,000	−0.0	0.91	96	+0.8	0.84	98	+2.5	0.98	94
			10,000	−0.1	1.02	95	+0.4	0.94	97	−0.1	0.92	98
Daily	0.720	0.119	1,000	+32.3	1.39	75	−12.1	1.33	74	+18.7	1.22	96
			10,000	+4.0	0.97	91	−1.3	0.99	93	+1.9	0.88	98

Unmarked exponential Hawkes process with stationary intensity one; time is measured in mean spacings between events. Intraday and Daily: the branching ratio and decay of the full-sample fits on Bitcoin of the intraday process of the Intersection events and of the daily process. 100 paths per row, fitted on $[0, T]$ with the start plan of the application. Bias: mean estimate minus the true value, in percent of the true value. Ratio: standard deviation of the estimates across paths over the asymptotic one. Cov.: number of the 100 paths on which the interval of ± 1.96 standard errors contains the true value. On none of the 800 fits is the likelihood at the estimate below the likelihood at the true parameters, and at $T = 1{,}000$ a search started at the true parameters reaches the same optimum on 398 of 400 paths; on the other 2, of the intraday configuration, the start plan ends at a log-likelihood higher by at most 0.002. At $T = 10{,}000$ in the intraday configuration the two searches agree on 91 of 100 paths, and on the other 9 the start plan is higher by at most 0.035.

75 and 74 of 100 paths. Slow mean reversion and the empty-history likelihood applied to stationary paths are possible contributors, which this exercise does not separate. The intraday configuration shows no comparable distortion. These checks justify caution for the fitted daily process, rather than a blanket claim of estimator calibration.

B.6 The Test Statistics at Estimated Parameters

Proposition B.4 concerns the true compensator. What can be computed is the plug-in

$$\hat{\tau}_q = \Lambda^*\Big((t_{q-1}, t_q]\,; \hat{\theta}\Big), \tag{B.8}$$

where $\hat{\theta}$ is obtained from (2.25) on the realization that the fit is then tested against. The fitted distribution has thereby moved toward the empirical one, the distances are stochastically smaller, and the limit laws of Appendix B.4 no longer apply: the law of the statistics then depends in general on the model and on the true parameters (Durbin, 1973), and no table of critical values is available

for the processes of this thesis. The size of the effect is measured by referring the plug-in statistics to the critical values of Appendix B.4.

Table B.5: Plug-in goodness-of-fit statistics against their known-parameter critical values

			$\sqrt{n}\,D_n$		W^2		Q	
Configuration	η	β	95% value	Rejection	95% value	Rejection	95% value	Rejection
Grid	0.5	1	0.862	0.00	0.126	0.00	11.649	0.07
Grid	0.9	1	0.984	0.00	0.195	0.00	11.222	0.06
Intraday	0.176	65.0	0.998	0.00	0.175	0.00	10.426	0.04
Daily	0.720	0.119	1.062	0.01	0.195	0.00	10.736	0.05

Unmarked exponential Hawkes process with stationary intensity one; time is measured in mean spacings between events. Intraday and Daily: the branching ratio and decay of the full-sample fits on Bitcoin of the intraday process of the Intersection events and of the daily process. 100 paths at $T = 1{,}000$, each fitted by maximum likelihood, with the durations computed at the estimate. 95% value: the empirical 95% quantile of the statistic across paths. Rejection: share of paths on which the statistic exceeds the known-parameter 5% value (1.358, 0.461 and 11.07).

Table B.5 shows the effect. The 95% value of $\sqrt{n}\,D_n$ at estimated parameters lies between 0.86 and 1.06, against 1.358, and that of W^2 between 0.13 and 0.20, against 0.461. Referred to the tabulated values, the two statistics reject a true model on at most one path in a hundred. The shift differs from one configuration to another, so no single corrected table would serve. The Ljung–Box statistic is not shifted.

The remedy is a reference distribution that reproduces the estimation step. For a statistic $\mathcal{D}$ measuring distance from Exp(1), Algorithm 3 simulates a replicate from the fitted process, re-estimates on that replicate, computes the statistic from the plug-in durations of the replicate, and reads the observed value against the resulting distribution.

With $\mathcal{D}_{\text{obs}}$ and $\mathcal{D}_1, \ldots, \mathcal{D}_B$ as in Algorithm 3, the p-value is

$$\hat{p} = \frac{1 + \sum_{b=1}^{B} I\{\mathcal{D}_b \geq \mathcal{D}_{\text{obs}}\}}{B + 1}. \tag{B.9}$$

Every replicate is re-estimated, which is what reproduces the effect of the estimation step. When the observed statistic and the replicates are exchangeable, the rank of the observed one among the $B+1$ values is uniform, so $\mathbb{P}(\hat{p} \leq a) = a$ at every level a for which $a(B+1)$ is an integer. Exchangeability holds exactly when the replicates are simulated at the true parameters. They are simulated at the estimate, and the exercise measures how far the property then holds: a true model is simulated, fitted and tested by Algorithm 3, with $B = 39$, and the rejection rates at 5% and 10% are reported.

Algorithm 3 Parametric bootstrap p-value for a distance statistic on the rescaled durations

Require: events $\{t_q\}_{q=1}^{n}$ on $[0,T]$; a parametric family with intensity $\lambda^*(\cdot\,;\theta)$; a distance statistic $\mathcal{D}$; replicate count B

Ensure: $\hat{p}$ for the Exp(1) hypothesis on the rescaled durations

$\hat{\theta} \leftarrow \arg\max_\theta \ell(\theta)$ ▷ (2.25) on the observed realization

$\hat{\tau}_q \leftarrow \Lambda^*\Big((t_{q-1}, t_q]\,;\hat{\theta}\Big), \quad q = 1,\dots,n$ ▷ (B.8)

$\mathcal{D}_{\text{obs}} \leftarrow \mathcal{D}(\hat{\tau}_1,\dots,\hat{\tau}_n)$

for $b \leftarrow 1$ **to** B **do**

  simulate $\{t_q^{(b)}\}_{q=1}^{n_b}$ on $[0,T]$ from $\lambda^*(\cdot\,;\hat{\theta})$ ▷ Algorithm 2, after burn-in

  $\hat{\theta}^{(b)} \leftarrow \arg\max_\theta \ell\Big(\theta; \{t_q^{(b)}\}\Big)$ ▷ re-estimated: this is what reproduces the bias

  $\hat{\tau}_q^{(b)} \leftarrow \Lambda^*\Big((t_{q-1}^{(b)}, t_q^{(b)}]\,;\hat{\theta}^{(b)}\Big)$

  $\mathcal{D}_b \leftarrow \mathcal{D}\Big(\hat{\tau}_1^{(b)},\dots,\hat{\tau}_{n_b}^{(b)}\Big)$

end for

return $\hat{p}$ of (B.9)

Table B.6: Rejection rates of the bootstrap goodness-of-fit test under a true model, in continuous time and on a time grid

		Continuous time		On the grid	
Configuration	Statistic	5%	10%	5%	10%
Intraday	$\sqrt{n}\,D_n$	0.06	0.09	0.08	0.13
	W^2	0.08	0.12	0.09	0.13
	Q	0.04	0.10	0.02	0.11
Daily	$\sqrt{n}\,D_n$	0.08	0.13	1.00	1.00
	W^2	0.06	0.14	1.00	1.00
	Q	0.06	0.08	0.35	0.47

Unmarked exponential Hawkes process with stationary intensity one; time is measured in mean spacings between events. Intraday and Daily: the branching ratio and decay of the full-sample fits on Bitcoin of the intraday process of the Intersection events and of the daily process. 100 paths at $T = 1{,}000$, each fitted and tested by the bootstrap with re-estimation, 39 replicates; a test rejects at level a when its p-value is at most a. On the grid: the event times of the same paths are moved to the end of their grid cell, at most one event per cell is kept, and the path is fitted and tested in the same way, with replicates in continuous time. The grid is the five-minute interval for the intraday configuration and the day for the daily one. No replicate failed.

The left half of Table B.6 reports the test in continuous time, on the intraday and the daily configuration. It rejects a true model at rates between 0.04 and 0.08 at the 5% level and between 0.08 and 0.14 at the 10% level, which 100 paths do not distinguish from the nominal values. This holds in the daily configuration although the estimator is biased there (Table B.4): the replicates reproduce the estimation step, its bias included.

Event Times on a Grid

The event times of Chapter 3 are not continuous: a jump day is an integer, and an intraday event is the end of a five-minute interval. The consequence for a test of fit can be bounded.

Proposition B.6 (A lower bound under a grid)**.** *Let the event times lie on a grid of spacing Δ, at most one per grid point, and let the durations be computed with any parameters of the family (2.17) with baseline μ. Then every duration after the first is at least $\mu\Delta$, and*

$$D_n \;\geq\; 1 - e^{-\mu\Delta} - \frac{1}{n}. \tag{B.10}$$

Proof. Two consecutive events are at least Δ apart and $\lambda^* \geq \mu$, so $\tau_q = \int_{t_{q-1}}^{t_q} \lambda^*(s)\,\mathrm{d}s \geq \mu\Delta$ for $q \geq 2$. The empirical distribution function of the durations is therefore at most $1/n$ below $\mu\Delta$, while the unit exponential has mass $1 - e^{-\mu\Delta}$ there. □

The bound does not shrink with the sample, so a correct model is rejected once the sample is large enough. Whether it matters depends on $\mu\Delta$, the expected number of baseline events in one grid cell. For a daily process with a baseline of about one event every ten days and a thousand events the bound alone exceeds the 5% critical value. For an intraday process on five-minute intervals $\mu\Delta$ is of the order of 10^{-3} and the bound is negligible. The bound concerns the short durations that the grid removes and not the remaining effect of rounding, so the rejection rate is also measured, on paths simulated from a true model whose times are placed on each of the two grids and tested by Algorithm 3 with replicates in continuous time.

The right half of Table B.6 and Table B.7 report the result. On the daily grid the two distance statistics reject the true model on every one of the 100 paths. The grid also changes the estimate: it removes 17% of the events, and the fitted branching ratio falls from 0.72 to a median of 0.56 and the fitted decay from

Table B.7: A true model observed on a time grid: events lost, fitted parameters and the Kolmogorov–Smirnov statistic

			Events	η		β		$\sqrt{n}\,D_n$	
Configuration	Grid	Cell	lost (%)	True	Fitted	True	Fitted	Bound	Observed
Intraday	Five minutes	0.0016	1.0	0.176	0.170	65	59.8	0.01	0.64
Daily	One day	0.334	17.4	0.720	0.557	0.119	0.0775	3.30	5.59

The gridded paths of Table B.6. Cell: the grid spacing in mean spacings between events. Fitted: median estimate across the 100 gridded paths. Bound: median of $\sqrt{n}$ times the right-hand side of (B.10) at the fitted baseline; a dash where it is not positive. Observed: median statistic.

0.119 to 0.078. On the five-minute grid 1.0% of the events are lost, the fitted branching ratio is 0.170 against 0.176 and the fitted decay 59.8 against 65.0, and the Kolmogorov–Smirnov test rejects on 8 of 100 paths, against 6 in continuous time.

In this exercise a daily grid is sufficient for a correct model to be rejected. This is the reason for which the test of fit is applied to the intraday processes of Bitcoin, with replicates placed on the five-minute grid, and not to the daily processes (Section 3.3.1).

B.7 The Expected Count from a Given State

The conditional-count check starts the unmarked exponential simulator from a given intensity, without burn-in, and averages the future count over 20,000 independent draws. Table B.8 reports six states for the normalized intraday Bitcoin configuration. Time is measured in units of the mean spacing between events, about two days, so this exercise checks the formula and does not reproduce the one-day empirical forecast horizon.

The simulated means agree with the conditional-count formula: the largest discrepancy is 1.5 standard errors. The fitted slope and intercept across states also agree with the affine form, while the frozen-history approximation lies 20–22 standard errors below the means. Appendix B.8 extends the check to all four configurations and examines the estimated-parameter count. These finite exercises validate the implementation on the stated cases, not every marked or two-component specification.

Table B.8: The expected count against the mean of simulated counts, from six states at the origin

State	λ_t	Expected count	Simulated mean	Std. error	Difference	Frozen history
0	0.824	0.9967	0.9975	0.0085	+0.09	0.8245
0.5	0.912	0.9984	0.9964	0.0085	−0.23	0.8258
1	1.000	1.0000	1.0129	0.0085	+1.52	0.8272
2	1.176	1.0033	1.0028	0.0086	−0.06	0.8299
4	1.527	1.0098	1.0171	0.0086	+0.84	0.8353
8	2.229	1.0229	1.0206	0.0086	−0.28	0.8461

Unmarked exponential Hawkes process with $\eta = 0.176$, $\beta = 65.0$ and stationary intensity one: the branching ratio and the decay of the full-sample fit on Bitcoin of the intraday process of the Intersection events, in units of its mean spacing between events. State: the excitation $\lambda_t - \mu$ at the origin, in units of its stationary mean. Expected count: (2.28) over one unit of time. Simulated mean: average of 20,000 counts simulated forward from the state, with no burn-in. Difference: simulated mean minus expected count, in standard errors. Frozen history: the count when the events of the window are not allowed to excite further ones. A weighted regression of the six simulated means on λ_t gives a slope of 0.0161 and an intercept of 0.9873, against $b = 0.0187$ and $a = 0.9813$ in (2.29).

B.8 Count Verification and Plug-In Uncertainty

Appendix B.7 describes the check of Section 2.3.1 and reports it on the intraday configuration: the simulated mean count from a given state against (2.28), the slope and the intercept of (2.29), and the count under a frozen history, (2.3.1), as a negative control. Table B.9 summarizes it on the four configurations.

Table B.9: The expected count against the mean of simulated counts: four configurations

			Largest	Slope			Intercept			Frozen
Configuration	η	β	difference	b	Fitted	Diff.	a	Fitted	Diff.	history
Grid	0.5	1	1.42	0.7869	0.7840	−0.76	0.2131	0.2197	+1.14	−17.4
Grid	0.9	1	1.53	0.9516	0.9515	−0.05	0.0484	0.0542	+1.65	−12.6
Intraday	0.176	65.0	1.52	0.0187	0.0161	−0.35	0.9813	0.9873	+0.60	−20.1
Daily	0.720	0.119	1.72	0.9835	0.9794	−1.55	0.0165	0.0202	+1.03	−3.3

Unmarked exponential Hawkes process with stationary intensity one; time is measured in mean spacings between events. Intraday and Daily: the branching ratio and decay of the full-sample fits on Bitcoin of the intraday process of the Intersection events and of the daily process. Six states at the origin per configuration, as in Table B.8, 20,000 simulated counts each. Largest difference: the largest absolute difference between the simulated mean and (2.28) over the six states, in standard errors. Slope and intercept: weighted regression of the simulated means on λ_t, against b and a of (2.29), with the difference in standard errors. Frozen history: the count with the history frozen at the origin minus the simulated mean, in standard errors, at the state where it is closest.

In the 24 cells the simulated mean lies within 1.72 standard errors of the expected count, and the slopes and intercepts within 1.65 of their values. The

frozen history is rejected in every cell. Its margin is smallest in the daily configuration, where the decay is slow and few of the events of one unit of time have had time to excite others.

The regressor of Chapter 3 is (2.28) at the estimate, $G(\hat{\theta}) \coloneqq \mathbb{E}\Lambda_t(h;\hat{\theta})$, where the history is fixed and λ_t is recomputed from it at $\hat{\theta}$. By (2.2.4) and a second-order expansion around θ°,

$$\begin{aligned} \operatorname{Var}\Big(G(\hat{\theta})\Big) &\approx \frac{1}{T}\,\nabla G^\top\,\mathcal{I}(\theta^\circ)^{-1}\,\nabla G, \\ \mathbb{E}\Big[G(\hat{\theta})\Big] - G(\theta^\circ) &\approx \nabla G^\top\,\mathbb{E}\Big[\hat{\theta}-\theta^\circ\Big] + \tfrac{1}{2}\operatorname{tr}\Big(\nabla^2 G\,\operatorname{Var}(\hat{\theta})\Big). \end{aligned} \tag{B.11}$$

Both terms of the bias are of order $1/T$, so the plug-in count is expected near the true value and not on it. Its error matters because it is an error in a regressor. If the regressor observed is $x+u$, with u uncorrelated with x and with the regression error, the least-squares slope converges to the true one multiplied by $\sigma_x^2/(\sigma_x^2+\sigma_u^2)$. The ratio σ_u^2/σ_x^2 is reported, with σ_u^2 the variance of the plug-in count across estimates at a fixed history and σ_x^2 the variance of the true count across the origins of a stationary path. It is an indication only, since in the application consecutive origins share almost the same estimation window.

Table B.10: Bias and dispersion of the expected count evaluated at estimated parameters: four configurations

Configuration	η	β	T	True count	Mean	Bias (%)	Std. dev.	Predicted	σ_u^2/σ_x^2
Grid	0.5	1	1,000	0.8987	0.9046	+0.7	0.0561	0.0629	0.021
			10,000	0.8987	0.9013	+0.3	0.0172	0.0196	0.002
Grid	0.9	1	1,000	0.1935	0.1964	+1.5	0.0196	0.0190	0.000
			10,000	0.1935	0.1936	+0.1	0.0061	0.0060	0.000
Intraday	0.176	65.0	1,000	0.9967	0.9982	+0.1	0.0341	0.0381	3.737
			10,000	0.9967	0.9967	−0.0	0.0118	0.0121	0.446
Daily	0.720	0.119	1,000	1.0187	0.9878	−3.0	0.0672	0.0693	0.042
			10,000	1.0187	1.0156	−0.3	0.0196	0.0230	0.004

Unmarked exponential Hawkes process with stationary intensity one; time is measured in mean spacings between events. Intraday and Daily: the branching ratio and decay of the full-sample fits on Bitcoin of the intraday process of the Intersection events and of the daily process. The expected count (2.28) over one unit of time, evaluated at 100 maximum likelihood estimates obtained from independent paths of T events, with the history at the origin held fixed. Mean, bias and standard deviation are across the 100 estimates. Predicted: the standard deviation given by the first line of (B.11). Last column: the variance across estimates over the variance of the true count across the origins of a stationary path. The fixed history is the one at which the true count is at its median along a stationary path.

Table B.10 reports the plug-in count at a history with median closing intensity. Its bias is at most 3.0% of the true count at 1,000 events and at most 0.3% at 10,000. Its standard deviation falls by a factor between 2.9 and 3.4 between

the two sample sizes, against $\sqrt{10} = 3.16$, and the first line of (B.11) predicts it to within 12% at 1,000 events. The ratio σ_u^2/σ_x^2 is below 0.05 in the two grid configurations and in the daily one. In the intraday configuration it is 3.74 at 1,000 events and 0.45 at 10,000. With a decay this fast the expected count of the next period is almost the same from every state: its standard deviation across the origins of a stationary path is 0.018, against a standard deviation of the plug-in count across estimates of 0.034 at 1,000 events. At a second history, the one at which the true count is at its 90th percentile, the figures agree to the third decimal. The ratio is an upper indication of the attenuation for a second reason besides the shared estimation windows: the error u is not independent across the rows of one regression. An error in $\hat{\mu}$ or in $\hat{\alpha}$ at a given decay moves the whole history of the regressor by one affine map, which Section 2.3.1 shows to leave the fitted values unchanged, and only the error in $\hat{\beta}$ alters the regressor. The exercise does not separate the two.

B.9 The Likelihood Ratio

For each nested comparison, fit the null and alternative to the same observation window and compute twice their log-likelihood difference. Simulate from the fitted null, refit both models to every replicate with the same constraints and starting plan, and compare the observed statistic with the empirical upper quantile. A grid-aware version places simulated times on the observation grid before refitting. Negative numerical likelihood ratios diagnose an unsuccessful nested optimization; they are not evidence favoring the restricted model. This procedure supplies the bootstrap reference used in Chapter 3.

The Poisson–Hawkes and one–two-exponential comparisons lack a simple reference law and are judged by the bootstrap of Appendix B.9, whose replicates are in continuous time while the events lie on five-minute intervals. The bootstrap is examined on a comparison whose reference is known: the squared-return mark against the unmarked process, on the intraday events of Bitcoin. Its null sets the exponent δ at zero, on the boundary of its range, and the reference is $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ (Self and Liang, 1987), so the share of zeros among the bootstrap statistics estimates the weight of the mixture directly.

On ten rolling windows drawn at random, the statistic is referred to three critical values at the 5% level: the asymptotic one, 2.71; the one of Appendix B.9; and the one of the same algorithm with every replicate placed on the five-minute grid before both models are refitted. Under the null the mark carries no

information, so the marks of a replicate are drawn from those of the window, independently of its event times. Two limits apply. The comparison is of decisions on data windows and does not measure the size of the test. And the reference case has every other parameter identified, whereas the test against the Poisson process has, in addition, a decay that is not identified under the null (Section 2.2.6), so agreement in the first case does not by itself carry over to the second.

Table B.11: The likelihood ratio of the squared-return mark on rolling windows: asymptotic, bootstrap and grid-aware bootstrap decisions

		Asymptotic	Bootstrap			Grid-aware bootstrap			Share of
Events	Statistic	decision	5% value	p	Decision	5% value	p	Decision	zeros
Windows drawn at random									
483	4.55	reject	2.35	0.020	reject	2.64	0.020	reject	0.46
496	3.23	reject	3.26	0.055	keep	2.62	0.035	reject	0.52
517	0.00	keep	2.52	0.535	keep	2.20	0.500	keep	0.49
509	0.00	keep	3.16	0.850	keep	3.31	0.845	keep	0.54
526	0.00	keep	3.26	0.880	keep	2.24	0.815	keep	0.48
530	0.00	keep	2.95	0.565	keep	3.55	0.550	keep	0.49
505	0.00	keep	2.64	0.545	keep	3.33	0.520	keep	0.53
497	0.00	keep	2.23	0.525	keep	2.82	0.570	keep	0.54
257	0.00	keep	2.86	0.860	keep	3.18	0.885	keep	0.48
240	0.00	keep	2.34	0.945	keep	2.80	0.940	keep	0.51
Mean			2.76			2.87			0.50
Std. dev.			0.39			0.46			
Supplementary windows									
474	3.05	reject	3.57	0.065	keep	2.79	0.045	reject	0.49
506	0.29	keep	1.90	0.265	keep	2.25	0.325	keep	0.54
508	0.00	keep	3.70	0.590	keep	3.07	0.520	keep	0.48
306	0.00	keep	3.56	0.605	keep	3.09	0.610	keep	0.51

Bitcoin, intraday Intersection events, rolling windows of 1,000 days, in chronological order. Statistic: likelihood ratio of the process with the squared-return mark against the unmarked process. Asymptotic decision: at the 5% value 2.71 of $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$. Bootstrap: 199 replicates simulated from the fitted unmarked process, with marks drawn from those of the window independently of the event times, both processes refitted. Grid-aware: the same replicates with each event moved to the end of its five-minute interval, at most one per interval. Share of zeros: share of replicate statistics equal to zero, averaged over the two bootstraps. Supplementary windows: four windows added because one of the ten drawn has a statistic within a factor of two of 2.71; they were chosen by their statistic and are not a random draw.

Table B.11 reports the result. On nine of the ten windows the three decisions coincide: one rejection and eight non-rejections. On the remaining window the statistic is 3.23, the asymptotic value and the bootstrap on the grid reject, and the bootstrap in continuous time does not, with a p-value of 0.055. The bootstrap 5% value averages 2.76 in continuous time and 2.87 on the grid, against 2.71, and 50% of the replicate statistics are zero, against one half. The bootstrap therefore reproduces the reference law in a case where that law is known, and placing the replicates on the five-minute grid does not change it beyond its Monte Carlo

error.

That error is not small: with 199 replicates the standard deviation of the 5% value over the ten windows is 0.39 in continuous time and 0.46 on the grid, so the bootstrap does not settle a statistic that lies within about one unit of its critical value. The statistics for which the thesis relies on the bootstrap alone are for the most part far from it (Section 3.3.1).

The likelihood ratios of the daily processes of Table 3.11 were also referred to replicates placed on the grid of days, on which a jump day holds at most one event: the decisions for the displayed daily specifications are unchanged. No known-null size experiment was completed for Poisson against Hawkes or one exponential against two. Agreement with the marked reference case therefore validates that particular calibration check and does not certify finite-sample size for the nonstandard tests.

B.10 Empirical Goodness of Fit on Bitcoin

The fit of the adopted process was tested with the parametric bootstrap of Algorithm 3 on the full sample of Bitcoin, 1,532 events, with 199 replicates placed on the five-minute grid as the data are. The one-exponential process is rejected by the three statistics, the Kolmogorov–Smirnov and the Cramér–von Mises distances and the Ljung–Box statistic of the rescaled durations. The two-exponential process halves the first distance, from 0.105 to 0.056, and is still rejected by the two distances, while the dependence of the durations no longer is ($p = 0.35$). Neither process is therefore a complete description of the jump times, and after the second exponential the serial dependence that the first leaves is no longer detected. For the daily processes the grid of days is by itself sufficient for a correct model to be rejected (Appendix B.6), and the test is not applied to them.

Appendix C

The S&P 500 Application

This appendix collects the S&P 500 application: data and session conventions, process fits, full-sample regressions, matched forecasts, the two literature configurations and all-return controls. Table C.1 distinguishes their samples and estimation schedules. Differences across configurations cannot be attributed to one design choice alone.

Table C.1: Designs underlying the main forecast comparisons

Feature	SPX Clements–Liao	SPX Qu et al.	Primary comparison
Frequency	10 min	5 min	5 min
Intraday events	Intersection	—	Intersection
Daily events / post-jump	BNS	ABD	BNS
Backbone	E / M-D / M-I	ABD HAR-CJ (+ RS)	Bipower HAR-CJ
Cascade	1/5/22	1/5/22	1/7/30
HAR window	Recursive	Rolling 1,749	Rolling 1,000
Process refit	Daily	I daily; others 22	Daily
HAR refit	Daily	Daily	Daily
Forecasts	1,000	1,944	BTC 1,981; SPX 2,664
Following jump days	157	996	BTC 541; SPX 886

All forecasts are one day ahead and use information through the preceding close. Intersection means ABD, Lee–Mykland and sequential BNS intraday flags jointly. Cascade lengths count retained days; on SPX they count sessions. E and M backbones are defined in the text. ETH/LTC use the primary design without LCI. The Qu process window has 2,000 sessions; its HAR window has 1,749. Daily primary refits include tuning of F. Robustness uses 22-day process/half-life refits with daily HAR estimation.

C.1 Data and Session Clock

The one-minute S&P 500 OHLCV series comes from the Mendeley Data source documented by Santamaria (2020) and is aggregated to the sampling interval of each design. The stored sample runs from 2 January 2004 to 18 October 2019.

Of 3,977 sessions, 3,944 meet the requirement that at least 80% of the nominal 78 five-minute bars be stored (Tables 3.1 and C.2).

The first return is $100 \log(P_{t,1}/O_{t,1})$, from the first bar's open to its close; subsequent returns are percentage log changes between consecutive closes. The overnight close-to-open change is excluded. The event clock concatenates trading hours, so closures add no elapsed time. Ten-minute returns sum pairs of five-minute returns. This clock differs from the UTC calendar-day clock of the cryptocurrencies.

Table C.2: The S&P 500 series and its event sets

	Value	
Panel A: the stored series		
Sessions stored	3,944	
Five-minute bars a session	78	
Missing bars (%)	0.015	
Sessions with a missing bar	16	
Zero returns (%)	0.56	
Forecast origins	2,664	
Panel B: event sets		
Detector	Events	Events a session
BNS jump sessions (daily)	1,285	0.33
Intersection	336	0.09
ABD	1,648	0.42
Lee–Mykland	1,333	0.34
s-BNS	9,118	2.31

S&P 500, trading sessions of five-minute bars; event counts over all stored sessions. BNS: daily ratio test at 1%. Intraday detectors on five-minute returns; the Intersection keeps the returns flagged by ABD, Lee–Mykland and sequential BNS (s-BNS) together.

C.2 Replication of the Literature on the S&P 500

The two antecedents of this thesis forecast the volatility of an equity index with the intensity of a Hawkes process fitted to its jumps: Clements and Liao (2017) on Dow Jones futures and the constituents of that index, and Qu et al. (2024) on the S&P 500. Both fit the process to jump days and use the intensity predicted for the next day, here DI. Clements and Liao (2017) also fit it to intraday jumps and use the intensity integrated over the day that has ended, here SI. This section replicates both regressors on the S&P 500. For Qu et al. (2024) it is their index over a sample close to theirs, which runs from January 2004 to December 2019

and here ends in October 2019. For Clements and Liao (2017) it is a market of the same kind as theirs and not their data. The replication uses the primary design of Table C.1: five-minute returns, the Intersection events, the bipower HAR-CJ benchmark and the 1/7/30 cascade. The ladders in the configurations of the two studies themselves are reported in Appendix C.

The Fitted Processes

Table C.3 reports the full-sample estimates of the two processes of the replication, the intraday process on the Intersection and the daily process, and of the intraday process on each of the three tests of the Intersection taken alone. They are obtained by (2.2.4), with standard errors from (2.2.4).

Table C.3: Full-sample maximum likelihood estimates of the unmarked exponential Hawkes processes: S&P 500

Asset	Process	Events	$\hat{\mu}$	$\hat{\alpha} = \hat{\eta}$	$\hat{\beta}$	Half-life	Log-likelihood
SPX	Intraday, Intersection	336	0.076 (0.005)	0.105 (0.022)	5.280 (1.301)	51 min	$-1,133.8$
SPX	Intraday, ABD	1,648	0.141 (0.028)	0.670 (0.071)	0.037 (0.009)	18.6 d	$-3,034.5$
SPX	Intraday, s-BNS	9,118	0.426 (0.013)	0.816 (0.010)	5.757 (0.115)	47 min	6,351.7
SPX	Intraday, Lee–Mykland	1,333	0.258 (0.009)	0.238 (0.016)	11.384 (1.030)	24 min	$-2,436.5$
SPX	Daily, BNS days	1,285	0.104 (0.025)	0.688 (0.079)	0.014 (0.003)	50.3 d	$-2,693.8$

Time is measured in trading sessions of 390 minutes, and a day is a session. Standard errors from the numerically computed Hessian in parentheses. The intraday processes are fitted to the five-minute returns flagged by the Intersection and by each of its three tests taken alone (ABD, sequential BNS, Lee–Mykland), and the daily process to the days flagged by the BNS test. The half-life is $\log 2/\hat{\beta}$, in minutes of trading time or in days.

The two processes of the replication describe different dynamics. The daily process has a branching ratio of 0.69 and a half-life of fifty sessions: about two thirds of the jump days are attributed to earlier jump days, and the excitation lasts for weeks. The intraday process is fitted to a sparse record of 336 events. Its branching ratio is 0.11 and its half-life is 51 minutes, so the excitation of an intraday jump has decayed by the next session. The information on the decay is of order η^2 (Appendix B.5), so a decay at so low a branching ratio is imprecisely estimated, which its standard error confirms.

The single-detector rows show how the fitted dynamics depend on the event definition; they are not additional estimates of one fixed set of arrivals.

Full-Sample Regressions

Table C.4 reports HAR-CJ and its extensions estimated once on the full sample. Each regressor is computed with the estimates of Table C.3, or with a half-life

tuned on the full sample, so the table describes the fit of the models and is not a forecast. The last column adds the all-return control of Section 2.3.2 and is read in Appendix C.6.

Table C.4: Full-sample estimates of the HAR-CJ regressions on the S&P 500

	HAR-CJ	DI	SI	EI	LCI	F	EWRV
Constant	−0.003 (−0.08)	−0.199* (−4.06)	−0.005 (−0.10)	−0.251* (−2.08)	+0.835* (+2.21)	−0.004 (−0.11)	+0.229* (+4.34)
$\log C_{t-1}$	+0.545* (+22.15)	+0.546* (+22.16)	+0.545* (+22.07)	+0.544* (+22.03)	+0.544* (+22.00)	+0.545* (+22.15)	+0.207* (+3.67)
$\log C^{(7)}_{t-1}$	+0.294* (+8.16)	+0.296* (+8.44)	+0.294* (+8.07)	+0.300* (+8.37)	+0.300* (+8.36)	+0.300* (+8.41)	+0.241* (+6.49)
$\log C^{(30)}_{t-1}$	+0.031 (+0.90)	+0.079* (+2.35)	+0.031 (+0.90)	+0.027 (+0.78)	+0.027 (+0.79)	+0.026 (+0.77)	+0.023 (+0.70)
$\log(1 + J_{t-1})$	−0.078 (−1.62)	−0.080 (−1.65)	−0.079 (−1.61)	−0.094 (−1.89)	−0.094 (−1.87)	−0.094 (−1.91)	−0.256* (−3.96)
$\log(1 + J^{(7)}_{t-1})$	+0.057 (+0.43)	+0.053 (+0.42)	+0.057 (+0.43)	+0.048 (+0.37)	+0.049 (+0.37)	+0.043 (+0.33)	+0.028 (+0.22)
$\log(1 + J^{(30)}_{t-1})$	+0.715* (+3.98)	+0.332 (+1.85)	+0.716* (+3.98)	+0.736* (+4.10)	+0.736* (+4.10)	+0.738* (+4.10)	+0.775* (+4.42)
Jump-arrival regressor	—	+0.923* (+5.74)	+0.019 (+0.06)	+2.941* (+2.28)	+0.339* (+2.21)	+0.733* (+2.66)	—
$\log EWRV_{t-1}$	—	—	—	—	—	—	+0.408* (+6.83)
Adjusted R^2	0.7394	0.7417	0.7394	0.7401	0.7400	0.7406	0.7442

Ordinary least squares on the full sample, $n = 3,664$ sessions; bipower components and the 1/7/30 cascade counted in sessions. Newey–West t statistics with 30 lags in parentheses; * significant at 5%. Each column after the first adds to HAR-CJ the regressor named in its heading, computed with the full-sample estimates of the point process or a half-life tuned on the full sample. The EI columns are estimated with the expected count of the next day; with the intensity at the close the slope and the constant are rescaled and the t statistic of the slope is the same.

First, the continuous cascade carries the regression: its daily and weekly terms are large and precisely estimated in every column. Second, of the three jump terms of HAR-CJ only the monthly one is significant, and it is no longer significant once DI is added: the daily intensity, with a memory of weeks, takes its place. Third, DI is the one regressor of the literature that enters, with a t statistic of 5.74; SI is not significant. The regressors of this thesis enter as well, with smaller statistics: 2.28 for EI, 2.21 for LCI and 2.66 for F. In sample, the S&P 500 therefore supports the daily process of the literature most clearly, and the intraday integral of the literature not at all.

Out-of-Sample Comparison

Table C.5 reports HAR-CJ and its extensions on 2,664 forecasts, 886 of which follow a BNS jump day. DI and SI are the regressors of the literature; EI and LCI are those of this thesis and F is the comparison introduced here; the three are read in Section 3.2.

No extension improves HAR-CJ significantly under QLIKE, MSE or MAE. The changes of MSE lie between +0.05% and +0.26%, and the adjusted p-values of the contrasts of EI and LCI with DI and SI under that loss are all 1.000. DI lowers HRMSE by 3.92% and is significantly ahead of EI and LCI under that loss; F lowers it by 2.26%, significantly at 5%. Following jump days one comparison is significant, and it is unfavorable: SI raises MAE by 0.51%.

Table C.5: Primary intraday comparison: S&P 500

	Overall				Following jump days			
Model	QLIKE	MSE	MAE	HRMSE	QLIKE	MSE	MAE	HRMSE
HAR-CJ	0.2180	0.3759	0.4813	0.9780	0.2285	0.3870	0.4830	0.9133
DI	+0.13	+0.05	−0.22	−3.92**	+0.36	+0.45	+0.13	+0.35
SI	+0.07	+0.13	+0.14	+0.13	+0.39	+0.55	+0.51*	+1.01
EI	+0.25	+0.10	−0.01	−0.32	+0.92	+0.58	+0.28	+2.27
LCI	+0.28	+0.13	+0.01	−0.31	+1.02	+0.69	+0.35	+2.47
F	+0.34	+0.26	+0.12	−2.26*	+0.22	+0.30	+0.20	+0.22

Benchmark rows give absolute mean losses; all other rows give percentage changes against their matching benchmark (negative favors the extension). * and **: two-sided DM–HLN tests, Holm-adjusted, at 5% and 1%. Holm family: five benchmark contrasts per loss and sample. $n = 2,664$ overall, 886 following BNS jump days. Five-minute intersection events, bipower HAR-CJ, 1/7/30 cascade, rolling 1,000-day estimation and daily process, half-life and regression refits. Each extension adds the indicated regressor to HAR-CJ.

The replication therefore reproduces the result of the literature in one respect, the gain of the daily intensity under HRMSE. In the configurations of the two studies the daily intensity gains under more losses. In the ten-minute design of Clements and Liao (2017), in the form that adds the regressor to a cascade of bipower variation (form E of Appendix C), the daily intensity improves HRMSE by 20.03% and also MSE and MAE, the intraday integral improves the same three losses by 0.6% to 2.6%, and in the five-minute design of Qu et al. (2024) the reported unmarked and marked intensity extensions improve MAE and HRMSE significantly (Tables C.6 and C.7). Those designs differ from the primary one in their windows, event definitions, backbones and forecast periods, so the difference is not attributed to any one of them. The replication is partial and configuration-dependent.

C.3 The Model Ladders

The point-process construction of Clements and Liao (2017) represents jump clustering through the unmarked exponential Hawkes intensity of Section 2.2.3. Daily models use jump days as events; intraday models use flagged bars. The fitted process produces a regressor, which is then evaluated by its contribution to the volatility forecast. Their data are Dow Jones futures and the thirty constituents of that index, sampled every ten minutes from 2000 to 2012; their configuration is applied here to the S&P 500 alone.

The single-index implementation distinguishes two regression forms, E and M, which differ in the benchmark and in how the regressor enters. The notation is that of (2.33) and (2.34): C_t is the continuous component of (2.5), S_{t-1} is a series known at the close of day $t-1$, and the superscripts (w) and (m) denote the weekly and monthly terms, here over 5 and 22 sessions. In form E, the encompassing form of Clements and Liao (2017), the benchmark is the cascade of the continuous component and an extension adds the cascade of its regressor,

$$\begin{aligned}\log RV_t = \phi_0 + \phi_d \log C_{t-1} + \phi_w \log C_{t-1}^{(w)} + \phi_m \log C_{t-1}^{(m)} \\ + \vartheta_d S_{t-1} + \vartheta_w S_{t-1}^{(w)} + \vartheta_m S_{t-1}^{(m)} + \varepsilon_t,\end{aligned} \tag{C.1}$$

with $\vartheta_d = \vartheta_w = \vartheta_m = 0$ in the benchmark. In form M the benchmark also carries the jump cascade and an extension adds the regressor as one term, which is (2.34) with the 1/5/22 cascade,

$$\begin{aligned}\log RV_t = \phi_0 + \phi_d \log C_{t-1} + \phi_w \log C_{t-1}^{(w)} + \phi_m \log C_{t-1}^{(m)} \\ + \gamma_d \log(1 + J_{t-1}) + \gamma_w \log(1 + J_{t-1}^{(w)}) + \gamma_m \log(1 + J_{t-1}^{(m)}) + \vartheta S_{t-1} + \varepsilon_t,\end{aligned} \tag{C.2}$$

with $\vartheta = 0$ in the benchmark. Form E therefore asks whether the regressor adds to a benchmark without a jump term, and form M whether it adds to one that already has it. In the daily benchmark of form M, M-D, J_t is that of (2.5) on BNS jump days and zero otherwise; in the intraday benchmark M-I it is the sum of squared intersection-flagged returns. The regressors are the predictable daily intensity, the preceding jump-day indicator, the intensity integrated over the session just ended, and the intraday event count. E and M therefore identify forecasting forms; marks are a separate modeling choice. Constituent-cojump regressors cannot be transferred with only the index series available. Estimation and numerical conventions follow the implementation used here; the full original specification is given in Clements and Liao (2017).

Qu et al. (2024) motivate the daily unmarked (I) and marked (MI) specifications reported here. The mark is MedRV from (2.18), scaled by the largest event-day MedRV in the fitting window. The components of the benchmark are those of the ABD test of Section 2.1.2, with $\mathcal{J}_t^{\mathrm{ABD}}$ its flagged returns,

$$J_t^{\mathrm{ABD}} = \sum_{i \in \mathcal{J}_t^{\mathrm{ABD}}} r_{t,i}^2, \qquad C_t^{\mathrm{ABD}} = \max\{0,\, RV_t - J_t^{\mathrm{ABD}}\}. \tag{C.3}$$

The daily intensity terms are added to HAR-CJ and to HAR-CJ-RS, which adds to (2.33) the logarithms of the positive and negative realized semivariances of

the previous day (Barndorff-Nielsen et al., 2008; Corsi and Renò, 2012). The purpose of the ladder is to assess the added forecasting value of this structure. Numerical settings and the original motivation for each specification are left to the cited study.

C.4 Ten-Minute Clements–Liao Configuration

The retained history of the S&P 500 contains 148 ten-minute Intersection events and 933 BNS jump days. Recursive estimation produces 1,000 forecasts from 16 October 2015 to 18 October 2019, including 157 targets following a BNS jump day. Only 24 intraday Intersection events occur in this forecast period. Table C.6 presents all eleven models and their three matching benchmarks.

In the recursive scheme the estimation window starts at a fixed date and grows with the forecast origin. The reading of the paired test as a comparison of achieved accuracy (Section 2.5.3) requires a window of fixed length, so the paired tests of this section are descriptive.

Table C.6: Ten-minute literature-based ladder on the S&P 500

	Overall				Following jump days			
Model	QLIKE	MSE	MAE	HRMSE	QLIKE	MSE	MAE	HRMSE
E benchmark	0.2592	0.4525	0.5383	1.6196	0.3179	0.4397	0.5144	1.2012
E: daily λ	−1.81	−3.18**	−2.66**	−20.03**	−2.56	−2.48	−2.09	−14.92**
E: daily indicator	−2.03	−2.71**	−2.15**	−14.86**	−4.44	−2.67	−1.56	−9.60**
E: intraday integral	−0.46	−0.70**	−0.62**	−2.58**	+0.55	−0.12	−0.33	−2.20*
E: intraday count	−0.43	−0.69**	−0.62**	−2.65**	+0.62	−0.03	−0.30	−1.92*
M-D benchmark	0.2556	0.4454	0.5330	1.5032	0.3133	0.4356	0.5093	1.1176
M-D: daily λ	−0.35	−1.46*	−1.46**	−12.73**	−0.34	−0.64	−0.88	−7.81**
M-D: daily indicator	−0.32	−0.28	−0.24	−1.77**	+0.21	+1.80	+1.48	+13.60**
M-I benchmark	0.2584	0.4505	0.5362	1.5955	0.3184	0.4382	0.5118	1.1645
M-I: intraday integral	+0.04	+0.03	−0.00	+0.03	+0.35	+0.73	+0.43	+4.36
M-I: intraday count	+0.03	+0.03	−0.02	+0.03	+0.41	+0.86	+0.43	+5.53

Benchmark rows give absolute mean losses; all other rows give percentage changes against their matching benchmark (negative favors the extension). * and **: two-sided DM–HLN tests, Holm-adjusted, at 5% and 1%. Holm family: eight extensions per loss and sample. $n = 1{,}000$ overall, 157 following BNS jump days. Recursive estimation; daily refits; 1/5/22 cascade. E and M denote forecasting forms, with three distinct benchmarks. Intraday events use the triple intersection.

Daily intensity reduces overall HRMSE by 20.03% in E and 12.73% in M-D, both significant after adjustment. The E extension also improves MSE and MAE significantly, and so does M-D. The paired intensity-versus-indicator comparisons favor daily intensity under HRMSE in both forms. In form E the intraday integral and the intraday count also improve MSE, MAE and HRMSE significantly, by

0.6% to 2.7%, far less than the daily intensity, and in form M neither improves any loss. The integral does not show an advantage over the count: intraday counts have a small, significant HRMSE advantage over the integral in the overall sample. The sparse intraday record limits how far this comparison can support conclusions about an arrival-time model. The evidence chiefly supports the daily summary in selected losses within this single-index design.

C.5 Five-Minute Qu Configuration

This configuration uses five-minute returns, ABD jump days and a 1/5/22 cascade. It has 1,944 forecasts from 3 January 2012 to 18 October 2019, with 996 following an ABD jump day. The HAR window contains 1,749 sessions and the Hawkes window 2,000. The refit schedule is daily for I and every 22 sessions for MI; all HAR regressions refit daily.

Table C.7: Five-minute daily intensity ladder on the S&P 500

	Overall				Following jump days			
Model	QLIKE	MSE	MAE	HRMSE	QLIKE	MSE	MAE	HRMSE
HAR-CJ	0.2289	0.3970	0.5008	1.1922	0.2487	0.4008	0.4929	0.9776
HAR-CJ-I	−2.57	−2.51*	−2.12**	−11.66**	−0.94	−0.80	−0.82	−2.63
HAR-CJ-MI	−2.42	−2.33*	−1.70**	−9.47**	−1.35	−1.01	−0.71	−0.85
HAR-CJ-RS	0.2253	0.3898	0.4958	1.1435	0.2453	0.3944	0.4904	1.0112
HAR-CJ-RS-I	−2.05	−1.90	−1.56**	−9.93**	−1.05	−0.91	−0.50	−4.11
HAR-CJ-RS-MI	−1.88	−1.76	−1.23*	−7.93**	−1.32	−1.07	−0.47	−2.37

Benchmark rows give absolute mean losses; all other rows give percentage changes against their matching benchmark (negative favors the extension). * and **: two-sided DM–HLN tests, Holm-adjusted, at 5% and 1%. Holm adjustment retains the original eight-extension family, including omitted models. $n = 1{,}944$ overall, 996 following ABD jump days. I refits daily; MI every 22 sessions; HAR regressions daily. ABD decomposition and 1/5/22 cascade.

All four displayed intensity extensions lower overall mean loss under each metric. All four improve MAE significantly, and all four improve HRMSE at the 1% level. MSE gains are significant for I and MI on HAR-CJ only; no QLIKE gain is significant. Following ABD jump days, all mean changes remain favorable but none is significant after adjustment. The results support selected daily intensity gains without identifying a universally preferred marked specification. The recorded Holm adjustment retains the original eight-extension testing family, including omitted sign-channel models. Comparisons between I and MI also incorporate their different refit schedules.

The two ladders establish that daily arrival summaries can contribute to their stated benchmarks on the S&P 500, particularly under MAE and HRMSE. Their designs differ in more than the sampling frequency (Appendix C.2), and the primary design of Chapter 3 holds the backbone, origins and evaluation fixed within each asset to compare the alternative summaries directly.

C.6 All-Return Controls

The controls were also run on the S&P 500, under the design of Appendix C.2 and with the decay measured in session time (Table C.8). EWRV lowers the MSE of HAR-CJ by 1.83%, about a seventh of the Bitcoin gain, and the realized variance of the last hour of the session lowers it by 1.93%. Over EWRV no jump regressor adds under QLIKE, MSE or MAE, and both controls encompass the forecasts of EI, LCI, F and DI. Under HRMSE DI lowers the loss of the EWRV benchmark by 4.77%, with an adjusted p-value below 0.001: a daily process with a memory of weeks and a measure of the last hour of all returns carry different information under that loss.

Table C.8: The all-return control on the S&P 500

Model	QLIKE	MSE	MAE	HRMSE
HAR-CJ	0.2180	0.3759	0.4813	0.9780
HAR-CJ-EWRV	-2.67^{*}	-1.83^{*}	-0.88^{*}	-3.10
HAR-CJ-RV60	-2.09^{*}	-1.93^{**}	-0.86^{*}	-4.86^{**}
HAR-CJ-EWRV	0.2122	0.3690	0.4771	0.9476
HAR-CJ-EWRV-DI	-0.29	-0.28	-0.55	-4.77^{**}
HAR-CJ-EWRV-EI	$+0.99$	$+0.46$	$+0.13$	-0.06
HAR-CJ-EWRV-LCI	$+0.96$	$+0.46$	$+0.13$	-0.04
HAR-CJ-EWRV-F	-0.47	-0.33	-0.23	-2.20

The first block is measured against HAR-CJ and the second against HAR-CJ-EWRV. Benchmark rows give mean losses; the other rows give percentage changes against the benchmark of their block (negative favors the extension). * and **: significant at 5% and 1% after the Holm adjustment. The adjustment is over the four losses of each pair. Design of Table C.5, 2,664 forecasts. EWRV and RV60, the realized variance of the last hour of the session, enter in logs; the decay of EWRV is measured in session time.

C.7 Sampling Sensitivity

The equity rows of Table 3.19 find no significant EI or LCI improvement under QLIKE, MSE or MAE on any reported arm. F improves HRMSE under five-minute ABD events. The ten-minute rolling arm is not reported because

too many windows fall below the minimum-events rule. These results reinforce the distinction between the primary equity comparison and the literature configurations.

Appendix D

Supplementary Cryptocurrency Evidence

This appendix contains cryptocurrency evidence that supplements the main comparisons: alternative full-sample regressions, confidence-set sensitivity, cumulative losses, richer-process summaries and additional control checks. The main cross-asset forecast panel, timing placebo and six robustness comparisons are in Chapter 3; the equity application is in Appendix C.

D.1 Alternative Full-Sample Regressions

Table D.1 adds each arrival regressor to the HAR-RV cascade over the full sample. EI, LCI and F remain significant on Bitcoin, with smaller statistics than beside HAR-CJ, while DI is not significant. The regressor histories use full-sample estimates, so these coefficients describe associations and cannot establish a forecast gain.

Table D.1: Full-sample estimates of the HAR-RV regressions on Bitcoin

	HAR-RV	DI	SI	EI	LCI	F	EWRV	EWRV, EI
Constant	+0.174* (+4.87)	+0.202* (+4.79)	+0.264* (+5.24)	−0.296 (−1.32)	+0.640* (+3.11)	+0.164* (+4.60)	+1.287* (+15.14)	+2.460* (+9.37)
$\log RV_{t-1}$	+0.425* (+19.65)	+0.425* (+19.63)	+0.431* (+19.97)	+0.425* (+19.75)	+0.425* (+19.77)	+0.423* (+19.79)	+0.190* (+7.75)	+0.163* (+6.37)
$\log RV_{t-1}^{(7)}$	+0.411* (+12.73)	+0.409* (+12.61)	+0.404* (+12.34)	+0.413* (+12.94)	+0.413* (+12.95)	+0.413* (+12.87)	+0.349* (+11.74)	+0.339* (+11.25)
$\log RV_{t-1}^{(30)}$	+0.068* (+2.11)	+0.071* (+2.08)	+0.072* (+2.19)	+0.067* (+2.10)	+0.066* (+2.09)	+0.068* (+2.14)	+0.048 (+1.71)	+0.049 (+1.69)
Jump-arrival regressor	—	−0.091 (−0.78)	−0.204* (−2.47)	+0.997* (+2.12)	+0.620* (+2.30)	+0.112* (+2.27)	—	−2.216* (−4.77)
$\log EWRV_{t-1}$	—	—	—	—	—	—	+0.293* (+14.05)	+0.327* (+15.06)
Adjusted R^2	0.5475	0.5474	0.5482	0.5481	0.5482	0.5481	0.5880	0.5909

Ordinary least squares on the full sample, $n = 2,981$ days; 1/7/30 cascade of realized variance. Newey–West t statistics with 30 lags in parentheses; * significant at 5%. Each column after the first adds to HAR-RV, in levels, the regressors named in its heading, computed with the full-sample estimates of the point process or a half-life tuned on the full sample. The EI columns are estimated with the expected count of the next day.

D.2 Regressor and Confidence-Set Sensitivity

Table D.2 reports DI and SI across alternative cryptocurrency designs. The principal Bitcoin exception is SI under one-minute sampling, where MSE and MAE improve. Table D.3 gives the confidence sets configuration by configuration; their shared observations preclude interpreting retention frequencies as independent replications.

Table D.2: The regressors of the literature against HAR-CJ: Bitcoin, by event set and sampling frequency

	QLIKE	MSE	MAE	HRMSE
DI: the daily intensity, fitted to the BNS jump days				
Five-minute returns, daily refits	+0.06	−0.06	−0.09	−0.48
Five-minute returns, 30-day refits	+0.10	−0.06	−0.10	−0.40
One-minute returns	+0.60*	+0.13	−0.08	−0.74
Ten-minute returns	−0.30	+0.02	+0.03	−0.89
SI: the intraday intensity integrated over the day, by event set				
Intersection, daily refits (main)	+0.13	+0.06	+0.07	−0.21
Intersection, 30-day refits	+0.13	+0.06	+0.07	−0.17
ABD	−0.22	+0.07	+0.03	−0.07
Lee–Mykland	−0.25	−0.94	−0.36	−2.78
s-BNS	+0.04	−0.08	−0.00	−0.08
Intersection, arm A (ABD and s-BNS at 2%)	+0.28	+0.04	+0.07	−0.14
Intersection, arm B (Lee–Mykland at 2%)	+0.18	+0.07	+0.08	−0.11
Intersection, one-minute returns	−0.53	−1.16*	−0.76*	+0.60
Intersection, ten-minute returns	−0.04	+0.17	+0.05	+1.06

Percentage change in mean loss against HAR-CJ (negative favors the regressor), whole forecast sample of 1,981 days. The one- and ten-minute rows rebuild the realized measures, the backbone and the events at that frequency, so their losses are comparable within the row only. Rows other than the first of each block hold the Hawkes fits for 30 days. * and **: significant at 5% and 1% in the two-sided DM–HLN test, Holm-adjusted over the four losses of the row.

Figure D.1 displays cumulative daily loss differences. Its endpoints reproduce the overall gains, while the paths show that those gains do not accumulate uniformly. This is a description of the evaluation period, not a rule for choosing the model.

D.3 Richer Processes and State Summaries

Table D.4 compares the two-exponential conditional count, its fast and slow intensity components and a regression entering both components separately. The table uses 30-day process refits. None improves significantly on EI, and some state decompositions worsen MSE or HRMSE. It therefore assesses the transformation of the fitted states as well as the kernel.

Table D.3: Model confidence sets across configurations: Bitcoin

	QLIKE		MSE		MAE		HRMSE	
Configuration	T_R	$T_{\max}$	T_R	$T_{\max}$	T_R	$T_{\max}$	T_R	$T_{\max}$
Main (daily refits)	BDSELF	BDSELF	...ELF	...ELF	...EL.	.D.ELF	....L.	....L.
Main, 30-day refits	BDSELF	BDSELF	...EL.	...EL.	.D.EL.	.D.ELF	....L.	....L.
Backbone HAR-RV	BDSELF	BDSELF	.DSELF	BDSELF	BDSEL.	BDSEL.	...ELF	...ELF
Backbone SHAR	BDSELF	BDSELF	.DSELF	.DSELF	BDSEL.	BDSELF	....LF	...ELF
Backbone LHAR	BDSELF	BDSELF	BDSELF	BDSELF	BDSEL.	BDSEL.	....L.	....L.
Backbone LSHAR-CJ	BDSELF	BDSELF	...EL.	...EL.	...EL.	...EL.	....L.	....L.
Events ABD	.....F	.....F	.....F	.....F	...ELF	.....F	.....F	.....F
Events Lee–Mykland	.....F	.....F	.....F	.....F	...ELF	...ELF	.....F	.....F
Events s-BNS	BDSELF	BDSELF	....L.	BDSELF	....L.	BDSELF	....LF	....LF
Events Union	.....F	.....F	....LF	...ELF	....LF	...ELF	....LF	BDSELF
Test levels, arm A	BDSELF	BDSELF	...ELF	...ELF	...ELF	...ELF	....LF	...ELF
Test levels, arm B	BDSELF	BDSELF	...ELF	...ELF	.D.ELF	.D.ELF	....L.	....L.
One-minute returns	B.SELF	..SELF	.....F	.....F	.....F	.....F	.....F	.DSELF
Ten-minute returns	.....F	.D.ELF	.....F	.....F	BDSELF	...ELF	.....F	.....F

Bitcoin, six-model pool: the configuration's backbone alone and with one jump regressor. Each cell marks, in the fixed order BDSELF (backbone, DI, SI, EI, LCI, F), the models retained at the 25% level ($p > 0.25$); a dot marks an eliminated model. Each row changes one choice of the main configuration. T_R: range statistic; $T_{\max}$: maximum statistic. Moving blocks of 12 days; 10,000 bootstrap replications in the two main rows, 2,000 elsewhere. 1,981 one-day forecasts in every configuration.

Table D.4: The two-exponential process as a regressor, in four forms: Bitcoin

	Against HAR-CJ				Against HAR-CJ-EI			
Regressor	QLIKE	MSE	MAE	HRMSE	QLIKE	MSE	MAE	HRMSE
EI	−0.02	−1.12**	−0.57	−4.70**	—	—	—	—
EIbi	+0.38	−0.66	−0.35	−4.32**	+0.40	+0.46	+0.22	+0.40
EIbi$^+$	−0.08	−0.47	−0.35	−2.27**	−0.07	+0.65*	+0.23	+2.55**
EIbi$^-$	+0.60	−0.43	−0.20	−2.45*	+0.61	+0.70	+0.37	+2.37**
2EIbi	+0.50	−0.61	−0.32	−3.71**	+0.52*	+0.51*	+0.26	+1.04

Percentage change in mean loss of HAR-CJ augmented with the regressor of the row (negative favors it); $n = 1,981$ forecasts. EI: expected count of the one-exponential process. EIbi: expected count of the two-exponential process; EIbi$^+$ and EIbi$^-$: its fast and slow components; 2EIbi: both components as two regressors. Bipower HAR-CJ, 1/7/30 cascade, rolling 1,000-day estimation, processes refitted every 30 days. * and **: two-sided DM–HLN tests, Holm-adjusted over the four losses of each pair, at 5% and 1%.

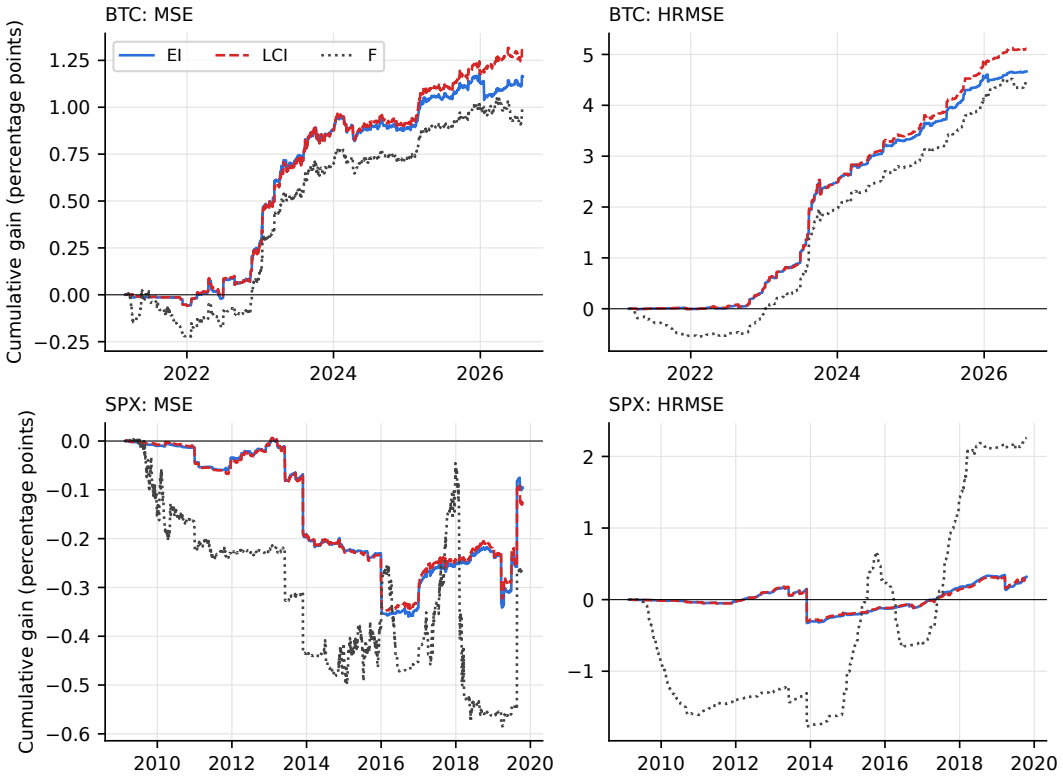


Figure D.1: Cumulative loss gains against HAR-CJ under the primary daily-refit design. Each curve is $100 \sum_{s \le t}(L_{0,s} - L_{m,s}) / \sum_{s=1}^{n} L_{0,s}$; its endpoint equals the percentage gain in the corresponding overall table. Positive values favor the extension. MSE uses log RV; HRMSE uses the squared relative level error. The curves describe when loss differences accrue; they are not sequential significance tests.

The additional daily-refit Bitcoin check also finds no whole-sample advantage from the two-exponential count: MSE is 0.50% above EI and HRMSE 0.35% above, without significance. Following jump days it is significantly worse under those two losses. Its rolling fast and slow half-lives have medians of 16 minutes and 1.4 days; these rolling summaries should not be equated with the full-sample timescales in Table 3.4.

Table D.5 extends the full-sample process comparison across event definitions. Parameter estimates and the fitted contribution of marks depend on which returns are flagged. These are additional descriptions of event fit and do not constitute further out-of-sample forecast tests.

D.4 Additional Control and Boundary Comparisons

Table D.6 retains the fuller ABD comparison, including F, and Table D.7 gives the cross-asset control results under shifted day boundaries. They supplement the compact Bitcoin tables in Chapter 3. Their loss-specific exceptions restrict the main MSE conclusion; they do not support a claim that every jump increment vanishes.

Table D.5: Full-sample Hawkes fits on every event set

Asset	Event set	Process	Events	Log-likelihood	Half-lives	η	δ	Optimum
Bitcoin	Intersection	One exponential	1,532	−2344.7	33 min	0.176	—	confirmed
		Two exponentials		−2214.2	22 min; 9.6 d	0.794	—	confirmed
		One exponential, mark		−2344.5	31 min	0.173	0.031	confirmed
		Two exponentials, mark		−2213.3	22 min; 10.7 d	0.803	0.051	confirmed
	ABD	One exponential	5,495	−1453.6	12 min	0.191	—	confirmed
		Two exponentials		−1401.9	12 min; 20.9 d	0.687	—	confirmed
		One exponential, mark		−1398.5	12 min	0.189	0.266	confirmed
		Two exponentials, mark		−1357.1	12 min; 48.7 d	0.635	0.219	confirmed
	Lee–Mykland	One exponential	6,118	332.0	16 min	0.312	—	confirmed
		Two exponentials		358.4	16 min; 43.0 d	0.696	—	confirmed
		One exponential, mark		339.0	15 min	0.310	0.071	confirmed
		Two exponentials, mark		365.4	15 min; 54.4 d	0.722	0.065	confirmed
	s-BNS	One exponential	7,689	13535.5	113 min	0.834	—	confirmed
		Two exponentials		13683.9	26 min; 223 min	0.852	—	confirmed
		One exponential, mark		13535.5	113 min	0.834	0.000	δ at 0
		Two exponentials, mark		13683.9	26 min; 223 min	0.852	0.000	δ at 0
	Union	One exponential	12,945	16240.6	42 min	0.608	—	confirmed
		Two exponentials		16621.6	21 min; 472 min	0.706	—	confirmed
		One exponential, mark		16240.6	42 min	0.608	0.000	δ at 0
		Two exponentials, mark		16621.6	21 min; 472 min	0.706	0.000	δ at 0
Ethereum	Intersection	One exponential	1,673	−2353.6	36 min	0.198	—	confirmed
		Two exponentials		−2229.9	24 min; 8.3 d	0.774	—	confirmed
		One exponential, mark		−2353.2	35 min	0.196	0.051	confirmed
		Two exponentials, mark		−2229.5	24 min; 8.5 d	0.776	0.032	confirmed
	ABD	One exponential	5,550	−1388.7	13 min	0.194	—	confirmed
		Two exponentials		−1299.0	13 min; 15.1 d	0.763	—	confirmed
		One exponential, mark		−1319.9	13 min	0.192	0.312	confirmed
		Two exponentials, mark		−1263.7	13 min; 20.4 d	0.751	0.179	confirmed
	Lee–Mykland	One exponential	5,973	7.3	17 min	0.301	—	confirmed
		Two exponentials		48.3	17 min; 25.9 d	0.743	—	confirmed
		One exponential, mark		28.0	16 min	0.298	0.137	confirmed
		Two exponentials, mark		57.7	16 min; 35.9 d	0.751	0.083	confirmed
	s-BNS	One exponential	8,780	16447.1	96 min	0.836	—	confirmed
		Two exponentials		16568.1	30 min; 194 min	0.854	—	confirmed
		One exponential, mark		16447.1	96 min	0.836	0.000	δ at 0
		Two exponentials, mark		16568.1	30 min; 194 min	0.854	0.000	δ at 0
Litecoin	Intersection	One exponential	1,409	−2196.9	26 min	0.168	—	confirmed
		Two exponentials		−2056.2	19 min; 14.0 d	0.844	—	confirmed
		One exponential, mark		−2196.7	26 min	0.167	0.035	confirmed
		Two exponentials, mark		−2056.2	19 min; 14.0 d	0.844	0.000	δ at 0
	ABD	One exponential	4,822	−1624.7	11 min	0.186	—	confirmed
		Two exponentials		−1546.1	11 min; 21.5 d	0.781	—	confirmed
		One exponential, mark		−1560.6	11 min	0.185	0.319	confirmed
		Two exponentials, mark		−1501.1	11 min; 37.8 d	0.771	0.243	confirmed
	Lee–Mykland	One exponential	5,045	−587.1	13 min	0.277	—	confirmed
		Two exponentials		−554.0	13 min; 30.0 d	0.729	—	confirmed
		One exponential, mark		−558.5	13 min	0.274	0.183	confirmed
		Two exponentials, mark		−533.8	13 min; 44.7 d	0.708	0.147	confirmed
	s-BNS	One exponential	6,127	8208.6	92 min	0.785	—	confirmed
		Two exponentials		8320.9	86 min; 34.8 d	0.894	—	confirmed
		One exponential, mark		8208.6	92 min	0.785	0.000	δ at 0
		Two exponentials, mark		8320.9	86 min; 34.8 d	0.894	0.000	δ at 0

Full-sample maximum likelihood of the intraday ladder on every event set of five-minute returns; whole sample of each asset. Notation and checks as in the main fit table; standard errors are on record. Log-likelihoods compare only within one asset and event set.

Table D.6: The two controls with the events of the ABD test: Bitcoin

	MSE			Whole sample		
	whole	after a jump day	other days	QLIKE	MAE	HRMSE
Against HAR-CJ						
EI	-1.33^{**}	-2.17	-1.03^{*}	-0.73	-0.86^{**}	-3.32^{**}
F	-2.67^{**}	-4.22^{*}	-2.12^{**}	-2.56^{*}	-1.37^{**}	-9.07^{**}
RV60	-11.85^{**}	-13.31^{**}	-11.35^{**}	-11.36^{**}	-4.69^{**}	-34.06^{**}
EWRV	-12.91^{**}	-13.72^{**}	-12.63^{**}	-11.52^{**}	-5.39^{**}	-37.94^{**}
RV60 against the jump regressor						
against EI	-10.67^{**}	-11.38^{**}	-10.42^{**}	-10.70^{**}	-3.86^{**}	-31.80^{**}
against F	-9.44^{**}	-9.49^{**}	-9.42^{**}	-9.03^{**}	-3.37^{**}	-27.49^{**}
A jump regressor added to a control						
EI to RV60	-0.07	-0.43	$+0.05$	-0.09	-0.05	-0.21
F to RV60	$+0.04$	-0.18	$+0.11$	$+0.15$	-0.06	$+0.05$
EI to EWRV	-0.17	-0.66	-0.00	-0.11	-0.11	-0.57^{*}
F to EWRV	-0.67	-0.59	-0.69	-0.15	-0.50	-4.92^{**}
EI to both	-0.19	-0.71	-0.02	-0.21	-0.10	-0.71^{**}

Percentage change in mean loss of the first model against the second (negative favors the first). After a jump day: the 541 forecasts that follow a day flagged by the BNS test; other days: the remaining 1,440; whole: 1,981. * and **: two-sided DM–HLN tests at 5% and 1%, Holm-adjusted over the four losses of each pair and sample. Events of the ABD test on five-minute returns; Hawkes fits and half-lives held for 30 days; 1/7/30.

Table D.7: The all-return control under a change of the day boundary

		EWRV against HAR-CJ				EI added to EWRV		F added to EWRV	
Asset	Day starts (UTC)	QLIKE	MSE	MAE	HRMSE	MSE	HRMSE	MSE	HRMSE
BTC	00:00	-11.55^{*}	-12.94^{**}	-5.41^{**}	-37.98^{**}	-0.17	-0.36	$+0.26$	-1.54
	06:00	-12.98^{**}	-17.62^{**}	-6.72^{**}	-53.60^{**}	-0.96	-6.78^{**}	-0.46	-6.76^{**}
	12:00	-11.75^{**}	-18.54^{**}	-7.95^{**}	-60.17^{**}	-0.69	-6.26^{*}	-0.31	-5.85
ETH	00:00	-6.83	-8.91^{**}	-4.32^{**}	-28.81^{**}	-0.32	-0.56	$+0.47$	$+4.63$
	06:00	-9.32^{*}	-12.25^{**}	-5.19^{**}	-38.72^{**}	-0.35	-3.41	-0.61	-2.32
	12:00	-10.32^{*}	-13.54^{**}	-6.36^{**}	-44.44^{**}	$+0.10$	-1.44	-0.11	$+2.85^{*}$
LTC	00:00	-11.63	-8.08^{**}	-3.85^{**}	-16.01^{**}	$+0.16$	-0.14	$+1.34$	$+1.16$
	06:00	-8.57	-8.99^{**}	-4.53^{**}	-21.60^{**}	-0.09	-0.96	$+0.56$	-0.88
	12:00	-4.26	-7.76^{**}	-4.30^{**}	-24.64^{**}	-0.13	-0.39	$+0.38$	-0.46

Percentage change in mean loss (negative favors the first model). Realized measures, events, fits and forecasts are rebuilt with days starting at the stated hour, so targets differ across rows and comparisons are paired within a row only. Daily refits, five-minute Intersection events, 1/7/30 cascade. * and **: two-sided DM–HLN tests at 5% and 1%, Holm-adjusted over the 36 tests of each asset.

Appendix E

Data and Computational Reproducibility

The public source series and coverage rules are described in Section 3.1.1 and appendix C.1. The implementation uses the `hfhawkes` estimators and stored forecast outputs; tables are generated from those outputs. For M bars per day the Lee–Mykland lookback is $\lceil\sqrt{252M}\rceil$; the intraday normalization is the median absolute return divided by its daily scale, estimated before the first forecast origin and normalized to mean square one. The regression sample starts after 280 retained days of initialization and lag construction. A process window requires at least 30 events, and comparisons retain common eligible origins. Multistart likelihood optimization uses the constraints and solver checks of Appendix B.5. Intraday marks are squared flagged returns and daily marks are MedRV, divided by their maximum in the estimation window; full-sample fits instead use the full-sample maximum. EIbi, MEI and MEIbi are the conditional next-day counts from the corresponding intraday processes, using empirical mark moments and a branching-ratio cap of 0.999 where required by the recorded count implementation; DMI uses the marked daily intensity. EIbi's fast and slow alternatives are intensity states, as defined in Definition A.1. States incorporate available events between refits, and all regression histories use the parameters available at their forecast origin. A negative numerical nested likelihood ratio remains a diagnostic failure and is not silently recoded as evidence for the simpler process. Paired one-day tests use zero autocovariance lags, longer-horizon tests use $h-1$, and full-sample coefficient statistics use 30 Newey–West lags. The confidence sets use moving blocks, while the paired-loss and portfolio sensitivities use stationary blocks. Neither resampling procedure repeats process fitting or tuning. The code

is available from the author on request.

References

Aalborg, Halvor Aarhus, Peter Molnár, and Jon Erik de Vries (2019). "What Can Explain the Price, Volatility and Trading Volume of Bitcoin?" In: *Finance Research Letters* 29, pp. 255–265.

Aït-Sahalia, Yacine, Julio Cacho-Diaz, and Roger J. A. Laeven (2015). "Modeling Financial Contagion Using Mutually Exciting Jump Processes". In: *Journal of Financial Economics* 117.3, pp. 585–606.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold (2007a). "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility". In: *Review of Economics and Statistics* 89.4, pp. 701–720.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens (2001a). "The Distribution of Realized Stock Return Volatility". In: *Journal of Financial Economics* 61.1, pp. 43–76.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys (2001b). "The Distribution of Realized Exchange Rate Volatility". In: *Journal of the American Statistical Association* 96.453, pp. 42–55.

— (2003). "Modeling and Forecasting Realized Volatility". In: *Econometrica* 71.2, pp. 579–625.

Andersen, Torben G., Tim Bollerslev, and Dobrislav Dobrev (2007b). "No-Arbitrage Semi-Martingale Restrictions for Continuous-Time Volatility Models Subject to Leverage Effects, Jumps and I.I.D. Noise: Theory and Testable Distributional Implications". In: *Journal of Econometrics* 138.1, pp. 125–180.

Andersen, Torben G., Tim Bollerslev, Per Frederiksen, and Morten Ørregaard Nielsen (2010). "Continuous-Time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns". In: *Journal of Applied Econometrics* 25.2, pp. 233–261.

Andersen, Torben G., Dobrislav Dobrev, and Ernst Schaumburg (2012). "Jump-Robust Volatility Estimation Using Nearest Neighbor Truncation". In: *Journal of Econometrics* 169.1, pp. 75–93.

Anderson, Theodore W. and Donald A. Darling (1952). “Asymptotic Theory of Certain “Goodness of Fit” Criteria Based on Stochastic Processes”. In: *Annals of Mathematical Statistics* 23.2, pp. 193–212.

Andreou, Elena (2016). “On the Use of High Frequency Measures of Volatility in MIDAS Regressions”. In: *Journal of Econometrics* 193.2, pp. 367–389.

Bacry, E., S. Delattre, M. Hoffmann, and J. F. Muzy (2013). “Some Limit Theorems for Hawkes Processes and Application to Financial Statistics”. In: *Stochastic Processes and their Applications* 123.7, pp. 2475–2499.

Bacry, Emmanuel, Iacopo Mastromatteo, and Jean-François Muzy (2015). “Hawkes Processes in Finance”. In: *Market Microstructure and Liquidity* 1.1, p. 1550005.

Barndorff-Nielsen, Ole E., Silja Kinnebrock, and Neil Shephard (2008). *Measuring Downside Risk: Realised Semivariance*. CREATES Research Paper 2008-42. Center for Research in Econometric Analysis of Time Series (CREATES), Aarhus University.

Barndorff-Nielsen, Ole E. and Neil Shephard (2004). “Power and Bipower Variation with Stochastic Volatility and Jumps”. In: *Journal of Financial Econometrics* 2.1, pp. 1–37.

— (2006). “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation”. In: *Journal of Financial Econometrics* 4.1, pp. 1–30.

Baur, Dirk G. and Thomas Dimpfl (2018). “Asymmetric Volatility in Cryptocurrencies”. In: *Economics Letters* 173, pp. 148–151.

Bergsli, Lykke Øverland, Andrea Falk Lind, Peter Molnár, and Michal Polasik (2022). “Forecasting Volatility of Bitcoin”. In: *Research in International Business and Finance* 59, p. 101540.

Bień-Barkowska, Katarzyna (2020). “Looking at Extremes without Going to Extremes: A New Self-Exciting Probability Model for Extreme Losses in Financial Markets”. In: *Entropy* 22.7, p. 789.

Bień-Barkowska, Katarzyna and Rodrigo Herrera (2026). “When to be Discrete: The Importance of Time Formulation in the Modeling of Extreme Events in Finance”. In: *International Journal of Forecasting* 42.1, pp. 61–84.

Binance (2026). *Spot Market Data: Klines (Candlestick) Endpoint, `GET /api/v3/klines`*. Binance Spot API documentation. BTCUSDT, ETHUSDT and LTCUSDT spot OHLCV bars, retrieved in monthly chunks; sample downloaded through 2026-07-31.

Bollerslev, Tim (1986). “Generalized Autoregressive Conditional Heteroskedasticity”. In: *Journal of Econometrics* 31.3, pp. 307–327.

Byrd, Richard H., Peihuang Lu, Jorge Nocedal, and Ciyou Zhu (1995). "A Limited Memory Algorithm for Bound Constrained Optimization". In: *SIAM Journal on Scientific Computing* 16.5, pp. 1190–1208.

Catania, Leopoldo and Mads Sandholdt (2019). "Bitcoin at High Frequency". In: *Journal of Risk and Financial Management* 12.1, p. 36.

Chassot, Jonathan and Francesco Audrino (2026). "HARd to Beat: The Overlooked Impact of Rolling Windows in the Era of Machine Learning". In: *International Journal of Forecasting* 42.2, pp. 330–343.

Chavez-Demoulin, Valérie, Anthony C. Davison, and Alexander J. McNeil (2005). "Estimating Value-at-Risk: A Point Process Approach". In: *Quantitative Finance* 5.2, pp. 227–234.

Christensen, Kim, Mathias Siggaard, and Bezirgen Veliyev (2023). "A Machine Learning Approach to Volatility Forecasting". In: *Journal of Financial Econometrics* 21.5, pp. 1680–1727.

Christoffersen, Peter F. (1998). "Evaluating Interval Forecasts". In: *International Economic Review* 39.4, pp. 841–862.

Clark, Todd E. and Kenneth D. West (2007). "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models". In: *Journal of Econometrics* 138.1, pp. 291–311.

Clements, Adam and Yin Liao (2017). "Forecasting the Variance of Stock Index Returns Using Jumps and Cojumps". In: *International Journal of Forecasting* 33.3, pp. 729–742.

Clements, Adam and Daniel P. A. Preve (2021). "A Practical Guide to Harnessing the HAR Volatility Model". In: *Journal of Banking & Finance* 133, p. 106285.

Cont, Rama (2001). "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues". In: *Quantitative Finance* 1.2, pp. 223–236.

Corsi, Fulvio (2009). "A Simple Approximate Long-Memory Model of Realized Volatility". In: *Journal of Financial Econometrics* 7.2, pp. 174–196.

Corsi, Fulvio, Davide Pirino, and Roberto Renò (2010). "Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting". In: *Journal of Econometrics* 159.2, pp. 276–288.

Corsi, Fulvio and Roberto Renò (2012). "Discrete-Time Volatility Forecasting with Persistent Leverage Effect and the Link with Continuous-Time Volatility Modeling". In: *Journal of Business & Economic Statistics* 30.3, pp. 368–380.

Daley, Daryl J. and David Vere-Jones (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. 2nd ed. Probability and Its Applications. New York: Springer.

Davies, Robert B. (1987). "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative". In: *Biometrika* 74.1, pp. 33–43.

Davison, Anthony C. and Richard L. Smith (1990). "Models for Exceedances Over High Thresholds". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 52.3, pp. 393–425.

Delbaen, Freddy and Walter Schachermayer (1994). "A General Version of the Fundamental Theorem of Asset Pricing". In: *Mathematische Annalen* 300.1, pp. 463–520.

Diebold, Francis X. and Roberto S. Mariano (1995). "Comparing Predictive Accuracy". In: *Journal of Business & Economic Statistics* 13.3, pp. 253–263.

Dumitru, Ana-Maria and Giovanni Urga (2012). "Identifying Jumps in Financial Assets: A Comparison Between Nonparametric Jump Tests". In: *Journal of Business & Economic Statistics* 30.2, pp. 242–255.

Durbin, James (1973). "Weak Convergence of the Sample Distribution Function When Parameters Are Estimated". In: *Annals of Statistics* 1.2, pp. 279–290.

Embrechts, Paul, Thomas Liniger, and Lu Lin (2011). "Multivariate Hawkes Processes: An Application to Financial Data". In: *Journal of Applied Probability* 48A, pp. 367–378.

Engle, Robert F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4, pp. 987–1007.

Errais, Eymen, Kay Giesecke, and Lisa R. Goldberg (2010). "Affine Point Processes and Portfolio Credit Risk". In: *SIAM Journal on Financial Mathematics* 1.1, pp. 642–665.

Filimonov, Vladimir and Didier Sornette (2012). "Quantifying Reflexivity in Financial Markets: Toward a Prediction of Flash Crashes". In: *Physical Review E* 85.5, p. 056108.

— (2015). "Apparent Criticality and Calibration Issues in the Hawkes Self-Excited Point Process Model: Application to High-Frequency Financial Data". In: *Quantitative Finance* 15.8, pp. 1293–1314.

Fuentes, Fernanda, Rodrigo Herrera, and Adam Clements (2018). "Modeling Extreme Risks in Commodities and Commodity Currencies". In: *Pacific-Basin Finance Journal* 51, pp. 108–120.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov (2006). "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies". In: *Journal of Econometrics* 131.1–2, pp. 59–95.

Giacomini, Raffaella and Halbert White (2006). "Tests of Conditional Predictive Ability". In: *Econometrica* 74.6, pp. 1545–1578.

Gkillas, Konstantinos, Rangan Gupta, and Christian Pierdzioch (2021). "Forecasting Realized Volatility of Bitcoin Returns: Tail Events and Asymmetric Loss". In: *European Journal of Finance* 27.16, pp. 1626–1644.

Hansen, Peter R. and Asger Lunde (2006). "Realized Variance and Market Microstructure Noise". In: *Journal of Business & Economic Statistics* 24.2, pp. 127–161.

Hansen, Peter R., Asger Lunde, and James M. Nason (2011). "The Model Confidence Set". In: *Econometrica* 79.2, pp. 453–497.

Hansen, Peter Reinhard, Chan Kim, and Wade Kimbrough (2024). "Periodicity in Cryptocurrency Volatility and Liquidity". In: *Journal of Financial Econometrics* 22.1, pp. 224–251.

Hardiman, Stephen J., Nicolas Bercot, and Jean-Philippe Bouchaud (2013). "Critical Reflexivity in Financial Markets: A Hawkes Process Analysis". In: *European Physical Journal B* 86.10, p. 442.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold (1997). "Testing the Equality of Prediction Mean Squared Errors". In: *International Journal of Forecasting* 13.2, pp. 281–291.

— (1998). "Tests for Forecast Encompassing". In: *Journal of Business & Economic Statistics* 16.2, pp. 254–259.

Hawkes, Alan G. (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". In: *Biometrika* 58.1, pp. 83–90.

Hawkes, Alan G. and David Oakes (1974). "A Cluster Process Representation of a Self-Exciting Process". In: *Journal of Applied Probability* 11.3, pp. 493–503.

Holm, Sture (1979). "A Simple Sequentially Rejective Multiple Test Procedure". In: *Scandinavian Journal of Statistics* 6.2. JSTOR stable 4615733, pp. 65–70.

Huang, Xin and George Tauchen (2005). "The Relative Contribution of Jumps to Total Price Variance". In: *Journal of Financial Econometrics* 3.4, pp. 456–499.

Jorion, Philippe (2001). *Value at Risk: The New Benchmark for Managing Financial Risk*. 2nd ed. New York: McGraw-Hill.

Kirchner, Matthias (2016). "Hawkes and INAR(∞) Processes". In: *Stochastic Processes and their Applications* 126.8, pp. 2494–2525.

Koenker, Roger and Gilbert Bassett (1978). "Regression Quantiles". In: *Econometrica* 46.1, pp. 33–50.

Koutmos, Dimitrios (2018). "Return and Volatility Spillovers among Cryptocurrencies". In: *Economics Letters* 173, pp. 122–127.

Kupiec, Paul H. (1995). “Techniques for Verifying the Accuracy of Risk Measurement Models”. In: *Journal of Derivatives* 3.2, pp. 73–84.

Kyriazis, Nikolaos A. (2019). “A Survey on Empirical Findings about Spillovers in Cryptocurrency Markets”. In: *Journal of Risk and Financial Management* 12.4, p. 170.

Lallouache, Mehdi and Damien Challet (2016). “The Limits of Statistical Significance of Hawkes Processes Fitted to Financial Data”. In: *Quantitative Finance* 16.1, pp. 1–11.

Lee, Kyungsub (2024). “Multi-Kernel Property in High-Frequency Price Dynamics under Hawkes Model”. In: *Studies in Nonlinear Dynamics & Econometrics* 28.4, pp. 605–624.

Lee, Suzanne S. and Per A. Mykland (2008). “Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics”. In: *Review of Financial Studies* 21.6, pp. 2535–2563.

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard (2015). “Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes”. In: *Journal of Econometrics* 187.1, pp. 293–311.

Ljung, Greta M. and George E. P. Box (1978). “On a Measure of Lack of Fit in Time Series Models”. In: *Biometrika* 65.2, pp. 297–303.

Markowitz, Harry (1952). “Portfolio Selection”. In: *Journal of Finance* 7.1, pp. 77–91.

Massey, Frank J. (1951). “The Kolmogorov–Smirnov Test for Goodness of Fit”. In: *Journal of the American Statistical Association* 46.253, pp. 68–78.

Mei, Hongyuan and Jason Eisner (2017). “The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process”. In: *Advances in Neural Information Processing Systems 30*, pp. 6754–6764.

Mohler, G. O., M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita (2011). “Self-Exciting Point Process Modeling of Crime”. In: *Journal of the American Statistical Association* 106.493, pp. 100–108.

Nakamoto, Satoshi (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. White paper.

Ogata, Yoshiko (1978). “The Asymptotic Behaviour of Maximum Likelihood Estimators for Stationary Point Processes”. In: *Annals of the Institute of Statistical Mathematics* 30.2, pp. 243–261.

— (1988). “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. In: *Journal of the American Statistical Association* 83.401, pp. 9–27.

Ogata, Yosihiko (1981). "On Lewis' Simulation Method for Point Processes". In: *IEEE Transactions on Information Theory* 27.1, pp. 23–30.

Patton, Andrew J. (2011). "Volatility Forecast Comparison Using Imperfect Volatility Proxies". In: *Journal of Econometrics* 160.1, pp. 246–256.

Patton, Andrew J. and Kevin Sheppard (2015). "Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility". In: *Review of Economics and Statistics* 97.3, pp. 683–697.

Patton, Andrew J. and Haozhe Zhang (2026). "Bespoke Realized Volatility: Tailored Measures of Risk for Volatility Prediction". In: *Journal of Econometrics* 254, p. 106122.

Pierre, Sébastien (2024). "Hawkes Process Modelling of Financial Jumps: A Volatility Forecasting Approach". PhD thesis. School of Mathematics, Cardiff University.

Politis, Dimitris N. and Joseph P. Romano (1994). "The Stationary Bootstrap". In: *Journal of the American Statistical Association* 89.428, pp. 1303–1313.

Qiu, Yue, Zongrun Wang, Tian Xie, and Xinyu Zhang (2021). "Forecasting Bitcoin Realized Volatility by Exploiting Measurement Error under Model Uncertainty". In: *Journal of Empirical Finance* 62, pp. 179–201.

Qu, Hui, Tianyang Wang, Peng Shangguan, and Mengying He (2024). "Revisiting the Puzzle of Jumps in Volatility Forecasting: The New Insights of High-Frequency Jump Intensity". In: *Journal of Futures Markets* 44.2, pp. 218–251.

Reinhart, Alex (2018). "A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications". In: *Statistical Science* 33.3, pp. 299–318.

Rizoiu, Marian-Andrei, Young Lee, Swapnil Mishra, and Lexing Xie (2017). "Hawkes Processes for Events in Social Media". In: *Frontiers of Multimedia Research*. Ed. by Shih-Fu Chang. Vol. 17. Association for Computing Machinery and Morgan & Claypool, pp. 191–218.

Santamaria, Cesar (2020). *S&P 500 Index*. Mendeley Data, V1. One-minute OHLCV bars for the S&P 500 cash index.

Self, Steven G. and Kung-Yee Liang (1987). "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions". In: *Journal of the American Statistical Association* 82.398, pp. 605–610.

Shen, Dehua, Andrew Urquhart, and Pengfei Wang (2020). "Forecasting the Volatility of Bitcoin: The Importance of Jumps and Structural Breaks". In: *European Financial Management* 26.5, pp. 1294–1323.

Slim, Skander, Ibrahim Tabche, Yosra Koubaa, Mohamed Osman, and Andreas Karathanasopoulos (2023). "Forecasting Realized Volatility of Bitcoin: The Informative Role of Price Duration". In: *Journal of Forecasting* 42.7, pp. 1909–1929.

Tomlinson, Matthew F., David Greenwood, and Marcin Mucha-Kruczyński (2024). "2T-POT Hawkes Model for Left- and Right-Tail Conditional Quantile Forecasts of Financial Log-Returns: Out-of-Sample Comparison of Conditional EVT Models". In: *International Journal of Forecasting* 40.1, pp. 324–347.

Wang, Jiqian, Feng Ma, M. I. M. Wahab, and Dengshi Huang (2021). "Forecasting China's Crude Oil Futures Volatility: The Role of the Jump, Jumps Intensity, and Leverage Effect". In: *Journal of Forecasting* 40.5, pp. 921–941.

Wang, Jying-Nan, Hung-Chun Liu, and Yuan-Teng Hsu (2020). "Time-of-Day Periodicities of Trading Volume and Volatility in Bitcoin Exchange: Does the Stock Market Matter?" In: *Finance Research Letters* 34, p. 101243.

Wang, Yajing, Fang Liang, Tianyi Wang, and Zhuo Huang (2020). "Does Measurement Error Matter in Volatility Forecasting? Empirical Evidence from the Chinese Stock Market". In: *Economic Modelling* 87, pp. 148–157.

Zhu, Ciyou, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal (1997). "Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization". In: *ACM Transactions on Mathematical Software* 23.4, pp. 550–560.