

Rough Heston model as the scaling limit of bivariate cumulative heavy-tailed INAR processes: Weak-error bounds and option pricing

Yingli Wang*

Zhenyu Cui[†]Lingjiong Zhu[‡]

December 3, 2025

Abstract

This paper links nearly unstable, heavy-tailed *bivariate cumulative* $\text{INAR}(\infty)$ processes to the rough Heston model via a discrete scaling limit, extending scaling-limit techniques beyond Hawkes processes and providing a microstructural mechanism for rough volatility and leverage effect. Computationally, we simulate the *approximating* $\text{INAR}(\infty)$ sequence rather than discretizing the Volterra SDE, and implement the long-memory convolution with a *divide-and-conquer FFT* (CDQ) that reuses past transforms, yielding an efficient Monte Carlo engine for *European options* and *path-dependent options* (Asian, lookback, barrier). We further derive finite-horizon *weak-error bounds* for option pricing under our microstructural approximation. Numerical experiments show tight confidence intervals with improved efficiency; as $\alpha \rightarrow 1$, results align with the classical Heston benchmark, where α is the roughness specification. Using the simulator, we also study the *implied-volatility surface*: the roughness specification ($\alpha < 1$) reproduces key empirical features—most notably the steep short-maturity ATM skew with power-law decay—whereas the classical model produces a much flatter skew.

Keywords: Rough volatility, $\text{INAR}(\infty)$, scaling limit, FFT-based simulation, implied volatility surface, weak-error bounds for option pricing.

MSC: 60G22, 60H35, 91G20, 62M10, 60F17.

*School of Mathematics, Shanghai University of Finance and Economics, Shanghai, People's Republic of China; naturesky1994@gmail.com.

[†]School of Business, Stevens Institute of Technology, Hoboken, NJ 07030, United States of America; zcui6@stevens.edu.

[‡]Corresponding author. Department of Mathematics, Florida State University, Tallahassee, Florida, United States of America; zhu@math.fsu.edu. All code (including FFT acceleration and option pricers) is available at <https://github.com/gagawjbytw/INAR-rough-Heston>.

1 Introduction

The rough Heston model proposed by [El Euch and Rosenbaum \(2019\)](#) is a one-dimensional stochastic volatility model in which the asset price S has the following dynamic:

$$dS_t = S_t \sqrt{V_t} dW_t,$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \gamma (\theta - V_s) ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \gamma \nu \sqrt{V_s} dB_s,$$

where γ, θ, ν, V_0 are positive constants, W and B are two standard Brownian motions with correlation ρ . The parameter $\alpha \in (1/2, 1)$ governs the smoothness of the volatility sample path. It is a modified version of the celebrated Heston model proposed by [Heston \(1993\)](#), motivated from empirical observations of roughness in volatility time series; see [Gatheral et al. \(2018\)](#). In addition to these empirical findings, early evidence on pricing under rough volatility and its implications for short-maturity skews is provided of [Bayer et al. \(2016\)](#), while [Keller-Ressel and Majid \(2020\)](#) establish a comparison principle that clarifies how rough versus classical Heston affects the volatility surface. [El Euch and Rosenbaum \(2019\)](#) establish the link between nearly unstable Hawkes processes and fractional volatility models and obtain the characteristic function of the log-price in the rough Heston model, which can be expressed in terms of the solution of a fractional Riccati equation. Their work is based on the theoretical results of the scaling limit of nearly unstable Hawkes processes studied by [Jaisson and Rosenbaum \(2015, 2016\)](#).

We introduce the *integer-valued autoregressive* (INAR(∞)) process. The model is conveniently defined using the *reproduction operator* (also known as the thinning operator), denoted by “ \circ ”. For a non-negative integer-valued random variable Y and a non-negative constant η , the operation is defined as:

$$\eta \circ Y := \sum_{j=1}^Y \xi_j,$$

where $(\xi_j)_{j \geq 1}$ are independent and identically distributed (i.i.d.) random variables, referred to as *offspring variables*. For the INAR(∞) process with Poisson offspring, which we consider here, these variables follow a Poisson distribution with mean η , i.e., $\xi_j \sim \text{Poisson}(\eta)$. It is assumed that a separate, independent family of such offspring variables is used for each application of the operator.

With this operator, an INAR(∞) process $(X_n)_{n \in \mathbb{Z}}$ is an integer-valued time series model that satisfies the following system of stochastic difference equations:

$$\epsilon_n = X_n - \sum_{k=1}^{\infty} \alpha_k \circ X_{n-k}, \quad n \in \mathbb{Z}.$$

Here, $(\epsilon_n)_{n \in \mathbb{Z}}$ forms the *immigration sequence*, consisting of i.i.d. random variables following a Poisson distribution with mean μ ; the constant μ is thus called the *immigration parameter*. The constants $\alpha_k \geq 0$ are the *reproduction coefficients* for each non-negative integer lag k . In the term $\alpha_k \circ X_{n-k}$, the operator implies a sum of X_{n-k} i.i.d. Poisson random variables, each with mean α_k . These offspring are assumed to be independent of the immigration sequence and all other random variables in the model’s history.

A cumulative INAR(∞) process (also known as a discrete Hawkes process) starts from time 1, i.e. $(N_n)_{n \geq 1}$, and is defined by

$$N_n := \sum_{s=1}^n X_s. \tag{1.1}$$

Hawkes processes, introduced by [Hawkes \(1971\)](#), are continuous-time self-exciting point processes widely used in various fields. Discrete-time analogs, such as cumulative INAR(∞) processes, provide similar modeling capabilities, particularly for count data observed at fixed time intervals. However, the scaling limits of heavy-tailed, unstable cumulative INAR(∞) processes differ considerably from those of continuous-time counterparts, which warrants dedicated theoretical investigations. Under certain conditions, the Poisson

autoregressive process can be viewed as an $\text{INAR}(\infty)$ process with Poisson offspring. For a comprehensive discussion of Poisson autoregressive models and their connections to INAR and Hawkes processes, please refer to [Fokianos \(2024\)](#) and [Huang and Khabou \(2023\)](#). The structural link between $\text{INAR}(\infty)$ time series and Hawkes point processes is well documented: [Kirchner \(2016\)](#) develops the $\text{INAR}(\infty)$ framework and, in the main theorem, constructs an $\text{INAR}(\infty)$ -based family of point processes that converges to a given Hawkes process, thereby formalizing the discrete-continuous connection. Alternatively, the $\text{INAR}(\infty)$ process can be equivalently defined via its conditional intensity expression, which highlights its similarity to the Hawkes process. Specifically, if we let an $\text{INAR}(\infty)$ process $(X_n)_{n \geq 1}$ start from time 1, it can also be defined by:

$$\lambda_n := \mu + \sum_{s=1}^{n-1} \alpha_{n-s} X_s, \quad (1.2)$$

where $\mu > 0$ is the immigration rate, and $(\alpha_n)_{n \geq 1} \in \ell^1$ represents the offspring distribution with $\alpha_n \geq 0$ for all $n \in \mathbb{N}$. Given the history \mathcal{F}_{n-1} , the count X_n follows a Poisson distribution with parameter λ_n , i.e.,

$$X_n \mid \mathcal{F}_{n-1} \sim \text{Poisson}(\lambda_n).$$

Discrete Hawkes processes have been studied in several applications. For example, [Xu et al. \(2022\)](#) proposed this model to study the deposit and withdrawal behaviors of money market accounts. [Wang et al. \(2024\)](#) prove that the fractional CIR process limit can also be obtained by scaling a sequence of heavy-tailed nearly unstable cumulative $\text{INAR}(\infty)$ processes by imitating the method of [Jaisson and Rosenbaum \(2016\)](#). Specifically, they study a heavy-tailed unstable $\text{INAR}(\infty)$ with

$$\alpha_n \sim \frac{K}{n^{1+\alpha}} \text{ as } n \rightarrow \infty,$$

where the constant $\alpha \in (\frac{1}{2}, 1)$. Inspired by the continuous-time case of [Jaisson and Rosenbaum \(2016\)](#), they consider a sequence of INAR processes $(X_n^\tau)_{n \in \mathbb{N}}$, indexed by a positive integer $\tau \in \{1, 2, 3, \dots\}$. The asymptotic setting is that this index tends to infinity, i.e., $\tau \rightarrow \infty$ along the integers. The corresponding sequence of the intensity parameters (λ_n^τ) is defined for $n \in \mathbb{N}$ by

$$\lambda_n^\tau := \mu^\tau + \sum_{s=1}^{n-1} \alpha_{n-s}^\tau X_s^\tau, \quad n \in \mathbb{N},$$

where μ^τ is a sequence of positive real numbers. In this paper, we will show that this method is useful in obtaining the limit as a rough Heston model.

The work of [Wang et al. \(2024\)](#) provides a foundational link by demonstrating that a sequence of univariate heavy-tailed $\text{INAR}(\infty)$ processes converges to a fractional CIR process, which precisely describes the dynamics of the volatility component in the rough Heston model. However, their analysis is confined to this one-dimensional volatility process and does not address the full machinery required for asset pricing, which necessitates a joint model for both volatility and the asset price itself.

This paper builds upon and significantly extends their findings. Our primary contribution is to generalize the framework from a univariate to a bivariate setting. We establish the joint convergence of two interacting $\text{INAR}(\infty)$ processes to the complete two-dimensional rough Heston model, thereby providing a discrete-time microstructural foundation for both the price and volatility dynamics via bivariate cumulative $\text{INAR}(\infty)$. Furthermore, and perhaps more importantly from a practical standpoint, we leverage this theoretical result to propose and validate a novel and efficient simulation-based numerical scheme for pricing a wide range of options under the rough Heston model. Our work thus bridges the gap between the microstructural theory for volatility and its practical application in financial derivatives pricing.

Simulation of the rough Heston model, in particular, recently has attracted significant attention in the literature. [Liang et al. \(2017\)](#) proposed an Euler-Maruyama method. [Abi Jaber and El Euch \(2019\)](#) introduced a multi-factor approximation approach, and [Ma and Wu \(2022\)](#) developed a fast Monte Carlo method. [Callegaro et al. \(2021\)](#) made a significant breakthrough in solving the fractional Riccati equation

arising in the rough Heston model, developing a hybrid numerical algorithm based on power series expansions that substantially outperforms traditional methods in both speed and stability. More recently, [Richard et al. \(2021\)](#) and [Richard et al. \(2023\)](#) made significant contributions by studying Euler-type discrete-time schemes for the rough Heston model. Their work rigorously proves the convergence of the discrete-time schemes to solutions of modified Volterra equations, which are shown to share the same unique solution as the initial equations. Notably, [Yang et al. \(2025\)](#) presented a highly efficient continuous-time Markov chain (CTMC) approximation method for pricing European and American options under the rough stochastic local volatility models, including as a special case the rough Heston model. In parallel, [Bayer and Breneis \(2024\)](#) develop weak simulation schemes tailored to rough Heston, providing another efficient route to European pricing that complements discretization- or microstructure-based approaches. While our focus is on pricing via a discrete microstructural approximation, related semimartingale-approximation ideas have been used to tackle control problems under rough Heston; see [Ma et al. \(2023\)](#), who reduce the utility-maximization with loss aversion to a tractable multi-factor SV control and solve it by a dual-control Monte Carlo scheme. Fast and reliable numerical methods for pricing and calibration in the rough Heston model have recently been investigated by [Boyarchenko et al. \(2025\)](#). They combine a modified fractional Adams scheme with a robust Fourier inversion technique (the SINH-CB method) and provide a detailed analysis of numerical errors and calibration pitfalls in rough volatility models.

The modeling of financial market volatility has long been a central topic in quantitative finance, with recent advances focusing on capturing both microstructural dynamics and macroscopic price behaviors. In this paper, we for the first time establish a theoretical connection between nearly unstable bivariate heavy-tailed INAR(∞) processes and the rough Heston model. This offers a novel framework for understanding not only the microstructural foundation of the rough Heston model, but also the efficient simulation of it.

The main contributions of the paper are:

1. **Microstructural limit to rough Heston model.** We prove joint convergence of a *bivariate* heavy-tailed, nearly-unstable cumulative INAR(∞) sequence to the rough Heston model (Theorem 2.10), thereby providing a discrete-time microstructural foundation for both price and variance dynamics and an explicit link between liquidity asymmetry and the leverage effect.
2. **FFT-accelerated INAR simulator.** Building on the discrete-time limit (indexed by τ), we propose an approximate Monte Carlo scheme for the rough Heston model through simulating the corresponding INAR(∞) process. A **divide-and-conquer (CDQ)** implementation using the **fast Fourier transform (FFT)** propagates the long-memory convolution in $\mathcal{O}(\tau \log^2 \tau)$ time per path, enabling fast pricing with large path counts while preserving accuracy. In spirit, our approach is complementary to weak simulation schemes for rough Heston (see [Bayer and Breneis \(2024\)](#)), but it leverages a discrete microstructural limit to handle path-dependent payoffs efficiently.
3. **Weak-error guarantees for option pricing.** We obtain finite-horizon weak-error bounds. Fix $\alpha \in (\frac{1}{2}, 1)$ and $K > 0$, let $C^{\text{INAR}, \tau}(T, K)$ denote the European call prices in the INAR approximation model, and let $C(T, K)$ be the corresponding prices in the limiting rough-Heston model. Then, for any $T > 0$, with a microscopic INAR layer, then for any $T \in (0, 1]$,

$$|C^{\text{INAR}, \tau}(T, K) - C(T, K)| \leq C_{T, K} \left(\tau^{\frac{1}{2} - \alpha} + \tau^{-\frac{1}{4}} \right) \quad (\text{Theorem 3.14}).$$

For any $T \in (0, 1]$,

$$|AA^{\text{INAR}, \tau}(T, K) - AA(T, K)| \leq C_{T, K} \left(\tau^{\frac{1}{2} - \alpha} + \tau^{-\frac{1}{4}} \right) \quad (\text{Theorem 3.18}),$$

$$|LB^{\text{INAR}, \tau}(T, K) - LB(T, K)| \leq C_{T, K} \left(\tau^{\frac{1}{4} - \frac{\alpha}{2}} + \tau^{-1/8} \sqrt{\log \tau} \right) \quad (\text{Theorem 3.21}).$$

4. **Comprehensive option-pricing engine.** The proposed method prices European and path-dependent options (arithmetic Asian, lookback, and barrier) within a single framework, delivering tight 95% confidence intervals across various strikes and payoffs.

5. **Consistency with the classical case.** The framework nests the classical Heston model when $\alpha = 1$. The numerical results match the closed-form solution of [Heston \(1993\)](#) for European payoffs, and the benchmark based on Euler discretization for path-dependent payoffs.
6. **Implied-volatility surface evidence.** Using the same simulator, we reproduce the hallmark features of rough volatility, namely *steep short-maturity at-the-money (ATM) skew and power-law decay in T* .

We use the simulator to study the implied-volatility surface, comparing the roughness specification ($\alpha < 1$) with the classical benchmark ($\alpha = 1$); see Figure 1. The roughness specification ($\alpha < 1$) reproduces key empirical features—most notably the steep short-maturity ATM skew with power-law decay—whereas the classical model produces a much flatter skew.

Roadmap. Section 2 develops the discrete-time microstructural foundation: we introduce the bivariate cumulative INAR(∞) model, state the assumptions and parameterization, derive the renewal representation, and prove the scaling limit to rough Heston (Theorem 2.10), together with the parameter/correlation mapping (see (2.11) and (2.14)). Section 4 presents the FFT-accelerated INAR simulator and a comprehensive numerical study covering European and path-dependent options, together with an implied-volatility surface diagnostic for the rough case $\alpha = 0.62$. Section 3 establishes finite-horizon *weak-error bounds* for option pricing that quantifies the bias of the INAR-based approximation and its Brownian prelimit, offering guidance on the choice of the discretization level τ . Section 4 provides numerical experiments for European options and path-dependent options such as Asian, lookback and barrier options that illustrates the efficiency of our proposed method. Section 5 concludes. Technical proofs and auxiliary results are collected in the Appendix.

2 From INAR(∞) Process to Rough Heston Model

2.1 Discrete-time microstructural foundation for the rough Heston model

Let $\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}, \alpha^{(4)}$ denote four positive ℓ^1 sequences. We consider a sequence of bivariate cumulative INAR(∞) processes $(N^{\tau,+}, N^{\tau,-})$, indexed by a positive integer $\tau \in \{1, 2, 3, \dots\}$ that tends to infinity. For $n \in \mathbb{N}$, their intensity is defined as

$$\begin{pmatrix} \lambda_n^{\tau,+} \\ \lambda_n^{\tau,-} \end{pmatrix} := \hat{\mu}_\tau(n) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_\tau \sum_{s=1}^{n-1} \begin{pmatrix} \alpha_{n-s}^{(1)} & \alpha_{n-s}^{(3)} \\ \alpha_{n-s}^{(2)} & \alpha_{n-s}^{(4)} \end{pmatrix} \begin{pmatrix} X_s^{\tau,+} \\ X_s^{\tau,-} \end{pmatrix},$$

where $\hat{\mu}_\tau(n)$ is the time-dependent baseline intensity, to be specified in Definition 2.3, and $a_\tau \in (0, 1)$ denotes the decay parameter. Based on this framework, we model the tick-by-tick transaction price P_n^τ as

$$P_n^\tau = N_n^{\tau,+} - N_n^{\tau,-}.$$

Next, we present a set of assumptions for our discrete model, analogous to the continuous-time setup of [El Euch and Rosenbaum \(2019\)](#).

Assumption 2.1 (Model Specifications). *Our model for the sequence of bivariate INAR(∞) processes is based on the following set of assumptions, which are analogous to those of [El Euch and Rosenbaum \(2019\)](#) for the continuous-time Hawkes process.*

(i) **(Stability and Near-Instability)** Let

$$A := \begin{pmatrix} \|\alpha^{(1)}\|_1 & \|\alpha^{(3)}\|_1 \\ \|\alpha^{(2)}\|_1 & \|\alpha^{(4)}\|_1 \end{pmatrix}, \quad \text{and} \quad \mathcal{S}(A) \text{ be its spectral radius.}$$

The coefficients $\alpha^{(i)}$ are independent of τ . We fix the normalization

$$\mathcal{S}(A) = 1.$$

Then, for each τ , $\mathcal{S}(a_\tau A) = a_\tau < 1$. For the asymptotic analysis we work in the near-unstable regime

$$a_\tau \uparrow 1, \quad 1 - a_\tau \sim \gamma \tau^{-\alpha} \quad (\alpha \in (0, 1), \gamma > 0), \quad \mu_\tau = \mu \tau^{\alpha-1},$$

as $\tau \rightarrow \infty$.

- (ii) **(No Statistical Arbitrage)** The influence kernels are balanced componentwise so that past trades do not induce a predictable drift asymmetry between buys and sells. Precisely,

$$\alpha_n^{(1)} + \alpha_n^{(3)} = \alpha_n^{(2)} + \alpha_n^{(4)} \quad \text{for all } n \geq 1.$$

Under the parameterization in Definition 2.3, this is ensured by the choice of the constant matrix χ whose two rows coincide, which in turn yields $\lambda_s^{\tau,+} \equiv \lambda_s^{\tau,-}$.

- (iii) **(Liquidity Asymmetry)** The cross-impact of trades is asymmetric. We assume that a past sell order has a stronger influence on future buy orders than a past buy order has on future sell orders:

$$\alpha^{(3)} = \beta \alpha^{(2)}, \quad \text{for some constant } \beta > 1.$$

This implies $\alpha^{(4)} = \alpha^{(1)} + (\beta - 1)\alpha^{(2)}$.

- (iv) **(Heavy-Tailed Kernels)** The influence of past trades decays slowly, following a power law. For some positive constant C , we assume:

$$n^\alpha \sum_{s=n}^{\infty} \left(\alpha_s^{(1)} + \beta \alpha_s^{(2)} \right) \rightarrow C, \quad \text{as } n \rightarrow \infty.$$

Remark 2.2 (Intuition behind the assumptions.). Let us provide some financial and mathematical intuition for these assumptions.

- **Assumption 2.1(i)** establishes the critical regime necessary for obtaining a non-trivial continuous-time limit. A strictly stable process would converge to a constant mean, while a truly unstable one would explode. The “near-instability” condition pushes the system to a critical point where, under appropriate scaling, the cumulative effects of self-excitation converge to a continuous diffusion process.
- **Assumption 2.1(ii)** is a symmetry condition that prevents the order flow from having a predictable drift. If this condition were violated, past trades would systematically generate more buy (or sell) pressure, creating a statistical arbitrage opportunity.
- **Assumption 2.1(iii)** introduces an asymmetry observed in real financial markets. The condition $\beta > 1$ models the empirical fact that sell-offs can trigger stronger buying rebounds than rallies trigger selling pressure. This asymmetry in the micro-structure is the source of the negative correlation between asset returns and volatility (the leverage effect) in the resulting macroscopic model.
- **Assumption 2.1(iv)** is the key ingredient in generating “rough” volatility. The slow, power-law decay of the kernels implies that the memory of past events persists for a long time. This long-range dependence on the microscopic order flow is precisely what aggregates to the Hurst parameter $H < 1/2$ (i.e., rough paths) in the macroscopic volatility process. The parameter α in the kernel’s tail directly governs the roughness of the limiting volatility.

To fully specify the model, we now define the functional forms of the parameters introduced in Assumption 2.1. Our parameterization is constructed to satisfy the near-instability and heavy-tailed conditions, closely following the framework of [El Euch and Rosenbaum \(2019\)](#).

Definition 2.3 (Model Parameterization). Let $\beta > 1$, $1/2 < \alpha < 1$, $\gamma > 0$, $\xi > 0$ and $\mu > 0$ be constants. The parameters of the INAR(∞) process sequence are specified as follows for each integer $\tau \in \{1, 2, 3, \dots\}$:

- **Decay and Intensity Parameters:** The decay parameter a_τ and the exogenous order intensity μ_τ are set to:

$$a_\tau := 1 - \gamma\tau^{-\alpha}, \quad \text{and} \quad \mu_\tau := \mu\tau^{\alpha-1}. \quad (2.1)$$

- **Influence Kernels:** The influence matrix kernel α_n^τ is given by the product of a scalar kernel φ_n^τ and a constant matrix χ :

$$\alpha_n^\tau := \varphi_n^\tau \chi, \quad \text{where} \quad \chi := \frac{1}{\beta+1} \begin{pmatrix} 1 & \beta \\ 1 & \beta \end{pmatrix}.$$

The scalar kernel φ_n^τ is a rescaled version of a base kernel φ_n :

$$\varphi_n^\tau := a_\tau \varphi_n,$$

where the base kernel $(\varphi_n)_{n \geq 1}$ is defined as:

$$\varphi_n := \begin{cases} 1 - \frac{1}{\Gamma(1-\alpha)}, & n = 1, \\ \frac{1}{\Gamma(1-\alpha)} \left(\frac{1}{(n-1)^\alpha} - \frac{1}{n^\alpha} \right), & n \geq 2. \end{cases}$$

- **Baseline Intensity:** The time-dependent baseline intensity $\hat{\mu}_\tau(n)$ is defined as:

$$\hat{\mu}_\tau(n) := \mu_\tau + \xi \mu_\tau \left(\frac{1}{1 - a_\tau} \left(1 - \sum_{s=1}^{n-1} \varphi_s^\tau \right) - \sum_{s=1}^{n-1} \varphi_s^\tau \right).$$

Remark 2.4. This baseline intensity closely resembles Definition 2.1 of [El Euch and Rosenbaum \(2019\)](#), with the notable exception of the sequence $(\varphi_n)_{n \geq 1}$. Here, $(\varphi_n)_{n \geq 1}$ is specifically constructed to satisfy the unstable condition

$$\sum_{n=1}^{\infty} \varphi_n = 1,$$

and the heavy-tailed behavior

$$n^\alpha \sum_{k=n+1}^{\infty} \varphi_k \rightarrow \frac{1}{\Gamma(1-\alpha)} \quad \text{as} \quad n \rightarrow \infty.$$

To analyze the limit, we first express the intensity $\lambda_n^{\tau,+}$ as the solution to a discrete renewal-type equation. We start from the original definition of the intensity vector λ_n^τ :

$$\begin{pmatrix} \lambda_n^{\tau,+} \\ \lambda_n^{\tau,-} \end{pmatrix} = \hat{\mu}_\tau(n) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \sum_{s=1}^{n-1} a_\tau \alpha_{n-s}^\tau \begin{pmatrix} X_s^{\tau,+} \\ X_s^{\tau,-} \end{pmatrix} = \hat{\mu}_\tau(n) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \sum_{s=1}^{n-1} \varphi_{n-s}^\tau \chi \begin{pmatrix} X_s^{\tau,+} \\ X_s^{\tau,-} \end{pmatrix},$$

where we used the parameterization from Definition 2.3. Let us focus on the first component, $\lambda_n^{\tau,+}$. Substituting the form of matrix χ , we get:

$$\lambda_n^{\tau,+} = \hat{\mu}_\tau(n) + \frac{1}{1+\beta} \sum_{s=1}^{n-1} \varphi_{n-s}^\tau (X_s^{\tau,+} + \beta X_s^{\tau,-}).$$

Now, let $\mathbf{M}_n^\tau = (M_n^{\tau,+}, M_n^{\tau,-})^\top = \mathbf{N}_n^\tau - \sum_{s=1}^n \lambda_s^\tau$ be the martingale associated with the counting process \mathbf{N}_n^τ . This means $X_s^{\tau,\pm} = \lambda_s^{\tau,\pm} + (M_s^{\tau,\pm} - M_{s-1}^{\tau,\pm})$. Substituting this back into the expression for $\lambda_n^{\tau,+}$, and noting that $\lambda_s^{\tau,+} = \lambda_s^{\tau,-}$ (due to the structure of χ), we obtain a discrete renewal equation for $\lambda_n^{\tau,+}$:

$$\begin{aligned} \lambda_n^{\tau,+} &= \hat{\mu}_\tau(n) + \frac{1}{1+\beta} \sum_{s=1}^{n-1} \varphi_{n-s}^\tau (\lambda_s^{\tau,+} + (M_s^{\tau,+} - M_{s-1}^{\tau,+}) + \beta \lambda_s^{\tau,-} + \beta (M_s^{\tau,-} - M_{s-1}^{\tau,-})) \\ &= \hat{\mu}_\tau(n) + \sum_{s=1}^{n-1} \varphi_{n-s}^\tau \lambda_s^{\tau,+} + \frac{1}{1+\beta} \sum_{s=1}^{n-1} \varphi_{n-s}^\tau ((M_s^{\tau,+} - M_{s-1}^{\tau,+}) + \beta (M_s^{\tau,-} - M_{s-1}^{\tau,-})). \end{aligned}$$

This is a standard linear discrete renewal equation. Its solution can be expressed using the discrete renewal kernel ψ^τ , which is defined as $\psi^\tau = \sum_{k=1}^{\infty} (\varphi^\tau)^{*k}$, where $*$ denotes the discrete convolution operation. Since $\|\varphi^\tau\|_1 = a_\tau < 1$, the series converges and ψ^τ is well-defined. As established in Lemma 4.1 of Wang et al. (2024), the solution is given by:

$$\lambda_n^{\tau,+} = \hat{\mu}_\tau(n) + \sum_{s=1}^{n-1} \psi_{n-s}^\tau \hat{\mu}_\tau(s) + \frac{1}{1+\beta} \sum_{s=1}^{n-1} \psi_{n-s}^\tau (M_s^{\tau,+} - M_{s-1}^{\tau,+} + \beta M_s^{\tau,-} - \beta M_{s-1}^{\tau,-}). \quad (2.2)$$

2.2 The rough limits of cumulative INAR(∞) processes

First, we consider how to choose the right scaling constant. We have chosen in (2.1) that $\mu_\tau = \mu\tau^{\alpha-1}$, where μ is some positive constant. To obtain a nondegenerate limit, similar to Wang et al. (2024), for $t \in [0, 1]$, we consider $(1 - a_\tau)\lambda_{[t\tau]}^{\tau,+}/\mu_\tau$, and define the renormalized intensity

$$C_t^\tau := \frac{1 - a_\tau}{\mu_\tau} \lambda_{[t\tau]}^{\tau,+}.$$

After some calculations, this can be re-written as

$$C_t^\tau = \frac{1 - a_\tau}{\mu_\tau} \hat{\mu}_\tau([t\tau]) + (1 - a_\tau) \sum_{s=1}^{[t\tau]-1} \psi_{[t\tau]-s}^\tau \frac{\hat{\mu}_\tau(s)}{\mu_\tau} + \kappa\tau(1 - a_\tau) \sum_{s=1}^{[t\tau]-1} \psi_{[t\tau]-s}^\tau \sqrt{C_{s/\tau}^\tau} (B_{(s+1)/\tau}^\tau - B_{s/\tau}^\tau), \quad (2.3)$$

where

$$\kappa := \sqrt{\frac{1 + \beta^2}{\gamma\mu(1 + \beta)^2}}, \quad (2.4)$$

and

$$B_t^\tau := \frac{1}{\sqrt{\tau}} \sum_{s=1}^{[t\tau]-1} \frac{M_s^{\tau,+} - M_{s-1}^{\tau,+} + \beta(M_s^{\tau,-} - M_{s-1}^{\tau,-})}{\sqrt{\lambda_s^{\tau,+} + \beta^2 \lambda_s^{\tau,-}}}, \quad (2.5)$$

which will converge weakly to a standard Brownian motion B as $\tau \rightarrow \infty$ by the martingale central limit theorem; see Wang et al. (2024). We first make the vector notation explicit. For $n \geq 0$, let

$$\mathbf{N}_n^\tau := \begin{pmatrix} N_n^{\tau,+} \\ N_n^{\tau,-} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\lambda}_n^\tau := \mathbb{E}[\mathbf{N}_n^\tau - \mathbf{N}_{n-1}^\tau | \mathcal{F}_{n-1}] = \begin{pmatrix} \lambda_n^{\tau,+} \\ \lambda_n^{\tau,-} \end{pmatrix},$$

where $(\mathcal{F}_n)_{n \geq 0}$ is the natural filtration, \mathbf{N}_n^τ is the bivariate counting process up to step n , and $\boldsymbol{\lambda}_n^\tau$ is its (predictable) intensity vector. With this notation, for $t \in [0, 1]$, we define the rescaled cumulative counts, cumulative intensities, and the associated normalized martingale as

$$\mathbf{Y}_t^\tau := \frac{1 - a_\tau}{\tau^\alpha \mu} \mathbf{N}_{[t\tau]}^\tau, \quad \boldsymbol{\Lambda}_t^\tau := \frac{1 - a_\tau}{\tau^\alpha \mu} \sum_{s=1}^{[t\tau]} \boldsymbol{\lambda}_s^\tau, \quad \mathbf{Z}_t^\tau := \sqrt{\frac{\tau^\alpha \mu}{1 - a_\tau}} (\mathbf{Y}_t^\tau - \boldsymbol{\Lambda}_t^\tau).$$

Before proving the main results, we establish a key technical result regarding the weak convergence of the discrete renewal kernel, which will be used repeatedly. This result is a discrete-time analog of Lemma 4.3 of Jaisson and Rosenbaum (2016).

Lemma 2.5 (Weak Convergence of the Renewal Kernel). *Let $(\psi_n^\tau)_{n \geq 1}$ be the sequence of discrete renewal kernels introduced in the text preceding Equation (2.2). Define a sequence of step functions $(\zeta^\tau(t))_{t \geq 0}$ by*

$$\zeta^\tau(t) := \frac{(1 - a_\tau)}{a_\tau} \tau \psi_{[t\tau]}^\tau.$$

As $\tau \rightarrow \infty$, the measure $\zeta^\tau(t)dt$ converges weakly to a probability measure on $[0, \infty)$ with density $f_{\alpha,\gamma}(t) = \gamma t^{\alpha-1} E_{\alpha,\alpha}(-\gamma t^\alpha)$, where $E_{\alpha,\alpha}$ is the Mittag-Leffler function.

Proof. We provide the proof in Appendix A.1. \square

Proposition 2.6. *Under the parameterization in Definition 2.3, the sequence $(\Lambda^\tau, \mathbf{Y}^\tau, \mathbf{Z}^\tau)$ is C -tight and*

$$\sup_{t \in [0,1]} \|\Lambda_t^\tau - \mathbf{Y}_t^\tau\|_1 \rightarrow 0$$

in probability as $\tau \rightarrow \infty$, where $\|\cdot\|_1$ represents ℓ^1 -norm on \mathbb{R}^2 . Moreover, if (\mathbf{Y}, \mathbf{Z}) is a possible limit point of $(\mathbf{Y}^\tau, \mathbf{Z}^\tau)$, then \mathbf{Z} is a continuous martingale with $[\mathbf{Z}, \mathbf{Z}] = \text{diag}(\mathbf{Y})$.

Proof. We provide the proof in Appendix A.2. \square

Since the components of (\mathbf{Y}^τ) and (Λ^τ) are pure jump processes, the classical Kolmogorov-Chentsov tightness criterion cannot be applied. Lemma 3.5 of Horst et al. (2023) gives a new criterion to verify the C -tightness for a sequence of càdlàg processes, which is useful in proving Proposition 2.6. We restate here Lemma 3.5 of Horst et al. (2023).

Lemma 2.7 (Lemma 3.5 of Horst et al. (2023)). *If $\sup_{n \geq 1} \mathbb{E} \left[\left| X_0^{(n)} \right|^\beta \right] < \infty$ for some $\beta > 0$, then the sequence $(X^{(n)})_{n \geq 1}$ is C -tight if for $\tau \geq 0$ and some constant $\theta > 2$, the following two results hold.*

(i)

$$\sup_{k=0,1,\dots,\lfloor \tau n^\theta \rfloor} \sup_{h \in [0,1/n^\theta]} \left| \Delta_h X_{k/n^\theta}^{(n)} \right| \rightarrow 0 \text{ in probability as } n \rightarrow \infty.$$

(ii) *There exists some constants $C > 0$, $p \geq 1$, $m \in \{1, 2, \dots\}$ and pairs $\{(a_i, b_i)\}_{i=1,\dots,m}$ satisfying*

$$a_i \geq 0, b_i > 0, \rho := \min_{1 \leq i \leq m} \{b_i + a_i/\theta\} > 1,$$

such that for all $n \geq 1$ and $h \in (0, 1)$,

$$\sup_{t \in [0,T]} \mathbb{E} \left[\left| \Delta_h X_t^{(n)} \right|^p \right] \leq C \cdot \sum_{i=1}^m \frac{h^{b_i}}{n^{a_i}}.$$

We will apply this lemma to establish the tightness of each component of our scaled processes. Specifically, to prove the tightness of the sequence $(Y_t^{\tau,+})_{\tau \geq 1}$, we will let the generic process $X^{(n)}$ in Lemma 2.7 correspond to our process $Y^{\tau,+}$, with the index n replaced by τ .

Before we prove the convergence of Y^τ and Z^τ , let us first prove the following lemma.

Lemma 2.8.

$$\sup_{t \in [0,1]} \|\Lambda_t^\tau - \mathbf{Y}_t^\tau\|_1 \rightarrow 0 \text{ in probability,}$$

as $\tau \rightarrow \infty$.

Proof. We provide the proof in Appendix A.3. \square

This implies the uniform convergence to zero in probability of $Y^{\tau,+} - \Lambda^{\tau,+}$.

Next, by noting that

$$\Lambda_t^{\tau,+} = \Lambda_t^{\tau,-},$$

we obtain

$$\sup_{t \in [0,1]} |Y_t^{\tau,+} - Y_t^{\tau,-}| \rightarrow 0,$$

as $\tau \rightarrow \infty$. Therefore, if a subsequence of $Y^{\tau,+}$ converges to some Y , then the corresponding subsequence of $Y^{\tau,-}$ converges to the same Y . We thus have the following proposition regarding the limit points of $Y^{\tau,+}$ and $Y^{\tau,-}$.

Proposition 2.9. *If (Y, Y, Z^+, Z^-) is a possible limit point for $(Y^{\tau,+}, Y^{\tau,-}, Z^{\tau,+}, Z^{\tau,-})$, then (Y_t, Z_t^+, Z_t^-) can be written as*

$$Y_t = \int_0^t v_s ds, \quad Z_t^+ = \int_0^t \sqrt{v_s} dB_s^1, \quad Z_t^- = \int_0^t \sqrt{v_s} dB_s^2,$$

where (B^1, B^2) is a bivariate Brownian motion and the rate process $v = (v_t)_{t \geq 0}$ is the solution of

$$v_t = \xi(1 - f_{\alpha,\gamma}(t)) + f_{\alpha,\gamma}(t) + \sqrt{\frac{1 + \beta^2}{\gamma\mu(1 + \beta)^2}} \int_0^t f_{\alpha,\gamma}(t - s) \sqrt{v_s} dW_s^V, \quad (2.6)$$

where

$$W^V := \frac{B^1 + \beta B^2}{\sqrt{1 + \beta^2}},$$

and for $t > 0$,

$$f_{\alpha,\gamma}(t) = \gamma t^{\alpha-1} E_{\alpha,\alpha}(-\gamma t^\alpha), \quad f_{\alpha,\gamma}(t) := \int_0^t f_{\alpha,\gamma}(u) du = 1 - E_\alpha(-\gamma t^\alpha),$$

and $f_{\alpha,\gamma}(0) = 0$. Here $E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}$ is the two-parameter Mittag-Leffler function (and $E_\alpha = E_{\alpha,1}$). Equivalently, $\mathcal{L}\{f_{\alpha,\gamma}\}(z) = \frac{\gamma}{\gamma + z^\alpha}$ for $z > 0$. Furthermore, for any $\epsilon > 0$, the process v has Hölder regularity $\alpha - \frac{1}{2} - \epsilon$.

Proof. We provide the proof in Appendix A.4. □

We then state the main result of the paper.

Theorem 2.10. *Under Assumption 2.1 and the model parameterization specified in Definition 2.3, as $\tau \rightarrow \infty$, the sequence of processes $(\mathbf{\Lambda}_t^\tau, \mathbf{Y}_t^\tau, \mathbf{Z}_t^\tau)_{t \in [0,1]}$ converges in law under the Skorokhod topology to $(\mathbf{\Lambda}, \mathbf{Y}, \mathbf{Z})$, where the limit process is described by a rate process $v = (v_t)_{t \geq 0}$, where:*

$$\mathbf{\Lambda}_t = \mathbf{Y}_t = \int_0^t v_s ds \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{Z}_t = \int_0^t \sqrt{v_s} \begin{pmatrix} dB_s^1 \\ dB_s^2 \end{pmatrix},$$

and the rate process v is the unique solution of the following rough stochastic differential equation:

$$v_t = \xi + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \gamma (1 - v_s) ds + \gamma \sqrt{\frac{1 + \beta^2}{\gamma\mu(1 + \beta)^2}} \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{v_s} dW_s^V, \quad (2.7)$$

where $W^V := \frac{B^1 + \beta B^2}{\sqrt{1 + \beta^2}}$, and (B^1, B^2) is a bivariate Brownian motion with independent components. Furthermore, for any $\epsilon > 0$, the process v has Hölder regularity $\alpha - 1/2 - \epsilon$.

Proof. The proof is technical and is deferred to Appendix A.5. It follows a standard two-step approach: we first establish the tightness of the sequence of processes in Proposition 2.6, and then we identify the unique limit point in Proposition 2.9. □

2.3 Parameter mapping and correlation for the macroscopic limit

Throughout the limit, we define the Brownian motions driving price and variance via the orthogonal decomposition

$$W_t^X := \frac{1}{\sqrt{2}} (B_t^1 - B_t^2), \quad (2.8)$$

$$W_t^V := \frac{B_t^1 + \beta B_t^2}{\sqrt{1 + \beta^2}}, \quad (2.9)$$

with (B^1, B^2) a two-dimensional Brownian motion with independent components. A direct calculation gives the correlation

$$d\langle W^X, W^V \rangle_t = \frac{1 - \beta}{\sqrt{2(1 + \beta^2)}} dt =: \rho dt, \quad (2.10)$$

where $\rho \in \left(-\frac{1}{\sqrt{2}}, 0\right)$ is consistent with the leverage effect induced by the liquidity asymmetry $\beta > 1$.

Moreover, moving from the rate process v (for intensities) to the variance process

$$V_t := \theta v_t,$$

rescales the drift to $\gamma(\theta - V_t)$ and the diffusion coefficient accordingly. Matching the diffusion scale obtained from the INAR martingale CLT with the standard rough-Heston notation yields the identity

$$\nu = \sqrt{\frac{\theta(1 + \beta^2)}{\gamma\mu(1 + \beta)^2}} \iff \mu = \frac{\theta(1 + \beta^2)}{\gamma\nu^2(1 + \beta)^2}. \quad (2.11)$$

We will henceforth *define* ν via (2.11), so that, with $V = \theta v$ and the correlation ρ in (2.10), the limiting variance dynamics takes the rough Heston form in (2.13).

Then we can also build microscopic processes converging to the log-price. For $\theta > 0$ and $t \in [0, 1]$, define the (already properly rescaled) microscopic log-price process $P^\tau = (P_t^\tau)_{t \in [0, 1]}$ as

$$P_t^\tau := \sqrt{\frac{\theta}{2}} \sqrt{\frac{1 - a_\tau}{\mu\tau^\alpha}} \left(N_{\lfloor t\tau \rfloor}^{\tau, +} - N_{\lfloor t\tau \rfloor}^{\tau, -} \right) - \frac{\theta}{2} \frac{1 - a_\tau}{\mu\tau^\alpha} N_{\lfloor t\tau \rfloor}^{\tau, +} = \sqrt{\frac{\theta}{2}} (Z_t^{\tau, +} - Z_t^{\tau, -}) - \frac{\theta}{2} Y_t^{\tau, +}.$$

From Theorem 2.10, $(P^\tau)_{t \in [0, 1]}$ converges in law (in the Skorokhod topology on $D([0, 1])$) to the macroscopic price process $(P_t)_{t \in [0, 1]}$ given by

$$P_t := \sqrt{\frac{\theta}{2}} \int_0^t \sqrt{v_s} (dB_s^1 - dB_s^2) - \frac{\theta}{2} \int_0^t v_s ds.$$

Let $V_t := \theta v_t$ and $W_t^X := \frac{1}{\sqrt{2}}(B_t^1 - B_t^2)$. It follows from the equation for v_t in (2.7) that V_t represents the variance process. Then we have the following theorem.

Theorem 2.11. *Under the same conditions as Theorem 2.10, as $\tau \rightarrow \infty$, the rescaled price processes $(P_t^\tau)_{t \in [0, 1]}$ converge in law (Skorokhod topology) to*

$$P_t = \int_0^t \sqrt{V_s} dW_s^X - \frac{1}{2} \int_0^t V_s ds, \quad (2.12)$$

where $V_t = \theta v_t$ is the unique solution of the rough Heston Volterra stochastic differential equation (SDE):

$$V_t = \theta\xi + \frac{\gamma}{\Gamma(\alpha)} \int_0^t (t - s)^{\alpha-1} (\theta - V_s) ds + \frac{\gamma\nu}{\Gamma(\alpha)} \int_0^t (t - s)^{\alpha-1} \sqrt{V_s} dW_s^V, \quad (2.13)$$

where ν is defined from the INAR parameters via (2.11) and (W^X, W^V) is a two-dimensional Brownian motion defined in (2.8)-(2.9) whose correlation is

$$d\langle W^X, W^V \rangle_t = \rho dt, \quad \rho = \frac{1 - \beta}{\sqrt{2(1 + \beta^2)}}. \quad (2.14)$$

Proof. We provide the proof in Appendix A.6. □

Remark 2.12. From (2.14) and the condition that $\beta > 1$, the correlation coefficient ρ between the two Brownian motions must lie in the interval $\left(-\frac{1}{\sqrt{2}}, 0\right)$, which is consistent with El Euch and Rosenbaum (2019).

3 Weak-Error Bounds for Option Pricing

Throughout this section and the appendices, all weak-error bounds are derived for a maturity $T \in (0, 1]$. This is a global normalization adopted for consistency with the foundational scaling limit framework in Section 2 (e.g., Theorem 2.10), which was established on the normalized time interval $(0, 1]$. All subsequent theorems and propositions in this paper (for European, Asian, and lookback options) implicitly operate under the assumption $T \in (0, 1]$.

We first summarize the notations that will be used in the rest of the paper in Table 1.

Symbol	Definition
<i>European call</i>	
$C(T, K)$	European call price under the limiting rough-Heston model.
$C^{\text{aux}, \tau}(T, K)$	European call price under the ζ^τ -Volterra prelimit model.
$C^{\text{INAR}, \tau}(T, K)$	European call price in the INAR approximation model.
<i>Arithmetic-average (Asian) call</i>	
$AA(T, K)$	Arithmetic-average (Asian) call price under the limiting rough-Heston model.
$AA^{\text{aux}, \tau}(T, K)$	Arithmetic-average (Asian) call price under the ζ^τ -Volterra prelimit model.
$AA^{\text{INAR}, \tau}(T, K)$	Arithmetic-average (Asian) call price in the INAR approximation model.
<i>Lookback call</i>	
$LB(T, K)$	Lookback call price under the limiting rough-Heston model.
$LB^{\text{aux}, \tau}(T, K)$	Lookback call price under the ζ^τ -Volterra prelimit model.
$LB^{\text{INAR}, \tau}(T, K)$	Lookback call price in the INAR approximation model.
<i>Characteristic functions</i>	
$\phi_T^\tau(z), \phi_T(z)$	Characteristic functions of the (log-)price at time T in the prelimit and limiting models, respectively. For $P_T = \log(S_T/K)$, we also write $\phi_{P,T}^\tau(z)$ and $\phi_{P,T}(z)$.

Table 1: Summary of Notations

3.1 Preliminaries

We decompose the total weak error into two parts:

- (i) the error between the INAR-based prelimit and an auxiliary Volterra SDE with τ -dependent kernel;
- (ii) the error between this auxiliary SDE and the rough-Heston limit.

Unless stated otherwise, we reuse the symbol C (possibly with a subscript) for positive finite constants whose values may change from line to line. Dependencies are indicated by subscripts: C depends only on the model parameters $(\alpha, \gamma, \nu, \theta, V_0, S_0, \beta)$ and on the (normalized) horizon; C_T may additionally depend on T ; $C_{T,K}$ may depend on (T, K) . All such constants are independent of τ .

Lemma 3.1 (Second-order Abelian bound for the chosen base kernel). *Let $\hat{\varphi}(s) = \sum_{n \geq 1} \varphi_n e^{-sn}$ with (φ_n) as in Definition 2.3:*

$$\varphi_1 = 1 - \frac{1}{\Gamma(1-\alpha)}, \quad \varphi_n = \frac{1}{\Gamma(1-\alpha)} \left(\frac{1}{(n-1)^\alpha} - \frac{1}{n^\alpha} \right), \quad n \geq 2, \quad \alpha \in (0, 1).$$

Then, as $s \downarrow 0$,

$$\hat{\varphi}(s) = 1 - s^\alpha + R(s), \quad |R(s)| \leq C_R s,$$

for some $C_R > 0$ independent of s . In particular, for any $\delta \in (0, 1 - \alpha]$,

$$|R(s)| \leq C_R s^{\alpha+\delta}, \quad \text{as } s \downarrow 0,$$

i.e., the second-order Abelian expansion holds with $\delta = 1 - \alpha$.

Proof. We provide the proof in Appendix B.1. □

Let $X_t := \log S_t$ be the log-price under the risk-neutral measure and maturity $T > 0$ be fixed. We work with an affine-Volterra (“rough Heston-type”) variance process with kernel $f_{\alpha, \gamma}$, $\alpha \in (\frac{1}{2}, 1)$. (see [Abi Jaber et al. \(2019\)](#) for the affine-Volterra framework and [El Euch and Rosenbaum \(2019\)](#) for the rough-Heston characteristic function). Denote by $F(t) := \int_0^t f_{\alpha, \gamma}(s) ds$ its cumulative kernel. For each refinement level $\tau \in [1, \infty)$ (equivalently, mesh size $h_\tau := 1/\tau \in (0, 1]$), let ζ^τ be a (causal) approximating kernel with cumulative $F^\tau(t) := \int_0^t \zeta^\tau(s) ds$. We work in the regime $\tau \rightarrow \infty$ (i.e., $h_\tau \downarrow 0$); all rates such as $\mathcal{O}(\tau^{\frac{1}{2}-\alpha})$ are to be understood in this limit.

We introduce the causal convolution

$$(f * g)(t) := \int_0^t f(t-s)g(s)ds, \quad t \in [0, T].$$

Let $(\kappa, \theta, \xi, \rho)$ be the Heston parameters (with ρ the correlation between W^X and W^V defined in (2.8) and (2.9)). For each admissible z in the vertical strip, the auxiliary function $\iota(\cdot, z)$ is the unique locally integrable solution to the (fractional) Riccati-Volterra equation (see ([Abi Jaber et al., 2019](#), Thm. B.1)):

$$\iota(t, z) = \int_0^t f_{\alpha, \gamma}(t-s) \left(\frac{1}{2}(z^2 - z) + (\rho\xi z - \kappa)\iota(s, z) + \frac{1}{2}\xi^2 (\iota(s, z))^2 \right) ds, \quad t \in [0, T], \quad (3.1)$$

and

$$\mathcal{A}(T, z) := (f_{\alpha, \gamma} * \iota(\cdot, z))(T), \quad \mathcal{B}(T, z) := \kappa\theta \int_0^T \iota(s, z)ds. \quad (3.2)$$

For the approximating model driven by ζ^τ , $\iota^\tau(\cdot, z)$ is defined by (see ([Abi Jaber et al., 2019](#), Thm. B.1))

$$\iota^\tau(t, z) = \int_0^t \zeta^\tau(t-s) \left(\frac{1}{2}(z^2 - z) + (\rho\xi z - \kappa)\iota^\tau(s, z) + \frac{1}{2}\xi^2 (\iota^\tau(s, z))^2 \right) ds, \quad (3.3)$$

and

$$\mathcal{A}^\tau(T, z) := (\zeta^\tau * \iota^\tau(\cdot, z))(T), \quad \mathcal{B}^\tau(T, z) := \kappa\theta \int_0^T \iota^\tau(s, z)ds. \quad (3.4)$$

For any complex number z in a vertical strip, write the characteristic functions of X_T and X_T^τ as

$$\phi_T(z) := \mathbb{E}[e^{zX_T}] = \exp(\mathcal{B}(T, z) + V_0\mathcal{A}(T, z)), \quad \phi_T^\tau(z) := \mathbb{E}[e^{zX_T^\tau}] = \exp(\mathcal{B}^\tau(T, z) + V_0\mathcal{A}^\tau(T, z)),$$

where the components $(\mathcal{A}, \mathcal{B})$ and $(\mathcal{A}^\tau, \mathcal{B}^\tau)$ are defined in (3.2) and (3.4). This exponential-affine representation follows from the affine-Volterra framework ([Abi Jaber et al., 2019](#), Thm. 4.3); in the rough-Heston case this recovers the fractional Riccati / characteristic-function (CF) representation of ([El Euch and Rosenbaum, 2019](#), Thm. 4.1, Eq. (4.6)).

Before we proceed, we recall the Volterra–Grönwall lemma from the literature (see e.g., ([Gripenberg et al., 1990](#), Lem. 8.2)).

Lemma 3.2 (Volterra–Grönwall). *Let $u : [0, T] \rightarrow [0, \infty)$ be measurable and suppose*

$$u(t) \leq a + \int_0^t k(t-s)u(s)ds, \quad t \in [0, T],$$

with $a \geq 0$ and a kernel $k \in L^1(0, T)$, $k \geq 0$. Then

$$u(t) \leq a + (r * a)(t) \leq a \exp(\|k\|_{L^1(0, T)}),$$

*where r denotes the (Volterra) resolvent of k , i.e. it is the unique $r \in L^1(0, T)$ solving $r = k + k * r$.*

To control the moment strip around $\operatorname{Re}(z) = \frac{1}{2}$ under the risk-neutral measure, we derive a Novikov-type threshold guaranteeing well-posedness and analyticity of the affine-Volterra characteristic exponent. We first re-write $X_t = \log P_t$ where P is defined in (2.12) and (2.13) as:

$$X_t = X_0 - \int_0^t \frac{1}{2} c_X^2 V_s ds + \int_0^t c_X \sqrt{V_s} dW_s^X, \quad (3.5)$$

$$V_t = V_0 + \int_0^t f_{\alpha, \gamma}(t-s) \left[(\theta - V_s) ds + c_V \sqrt{V_s} dW_s^V \right], \quad (3.6)$$

where the constants c_X and c_V are diffusion loading constants. To connect this general framework to the model derived in Theorem 2.11, we match the coefficients. Comparing the log-price dynamics (2.12) with the general form (3.5), we identify $c_X = 1$. For the variance process, assuming the kernel $f_{\alpha, \gamma}(t)$ in (3.6) corresponds to the fractional kernel $\frac{\gamma}{\Gamma(\alpha)} t^{\alpha-1}$ used in (2.13), matching the terms inside the stochastic convolution yields $c_V = \nu$.

We conduct the weak-error analysis using the model's "mild" (or resolvent) representation (3.6), which is based on the limiting Volterra kernel $f_{\alpha, \gamma}$.

Lemma 3.3 (Risk-neutral strip via a Novikov-type threshold). *Fix $T > 0$. Set*

$$\mathbf{G}(T) := \int_0^T (f_{\alpha, \gamma}(s))^2 ds, \quad \Phi(s) := \int_0^{T-s} f_{\alpha, \gamma}(u) du, \quad s \in [0, T].$$

Define

$$\Theta_T := \frac{2}{c_V^2 \|\Phi\|_{L^\infty(0, T)}^2} \quad (3.7)$$

so that in particular

$$\Theta_T \geq \frac{2}{c_V^2 T \mathbf{G}(T)}.$$

Then there exists a $\delta > 0$, given by

$$\delta = \frac{\sqrt{1 + 2\Theta_T/c_X^2} - 1}{4},$$

such that:

- (i) *for all $z \in \mathbb{C}$ with $\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$ the affine-Volterra Riccati system for $(\mathcal{B}(\cdot, z), \mathcal{A}(\cdot, z))$ is well posed on $[0, T]$;*
- (ii) *the map $z \mapsto \mathcal{B}(T, z) + V_0 \mathcal{A}(T, z)$ is analytic on the open strip $\{\frac{1}{2} - \delta < \operatorname{Re}(z) < \frac{1}{2} + \delta\}$ and*

$$\sup_{\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \operatorname{Re}(\mathcal{B}(T, z) + V_0 \mathcal{A}(T, z)) < \infty.$$

Proof. We provide the proof in Appendix B.2. □

Remark 3.4. In our model derived from Theorem 2.11 specified in (3.5)-(3.6), we have $c_X = 1$ and $c_V = \nu$. We can compute the strip half-width δ using the formula from Lemma 3.3. We have $\Theta_T = \frac{2}{c_V^2 \|\Phi\|_{L^\infty(0, T)}^2} = \frac{2}{\nu^2 \|\Phi\|_{L^\infty(0, T)}^2}$. Plugging this into the formula $\delta = \frac{\sqrt{1 + 2\Theta_T/c_X^2} - 1}{4}$ yields:

$$\delta = \frac{1}{4} \left(\sqrt{1 + \frac{2\Theta_T}{1^2}} - 1 \right) = \frac{1}{4} \left(\sqrt{1 + \frac{4}{\nu^2 \|\Phi\|_{L^\infty(0, T)}^2}} - 1 \right).$$

Building on Lemma 3.3, we show that the same risk-neutral strip and cumulant bounds persist uniformly along the kernel approximations $(\zeta^\tau)_\tau$, which will be needed to pass to the limit $\tau \rightarrow \infty$.

Proposition 3.5 (Uniform strip for kernel approximations). *Assume the kernel approximation $\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)} \rightarrow 0$ and $\|F^\tau - F\|_{L^\infty(0,T)} \rightarrow 0$ as $\tau \rightarrow \infty$, and, in addition, the uniform L^2 -energy bound*

$$\sup_{\tau \geq 1} G^\tau(T) = \sup_{\tau \geq 1} \int_0^T (\zeta^\tau(s))^2 ds < \infty. \quad (3.8)$$

Then the conclusions of Lemma 3.3 extend uniformly in τ to the Riccati systems (3.3) and (3.4) for $(\mathcal{A}^\tau(\cdot, z), \mathcal{B}^\tau(\cdot, z))$: there exists $\delta > 0$ (determined by $\Theta^ := \frac{2}{c_V^2 \sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0,T)}^2}$ via the formula in Lemma 3.3) and $C_T < \infty$, both independent of τ , such that*

$$\sup_{\tau \geq 1} \sup_{\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \operatorname{Re}(\mathcal{B}^\tau(T, z) + V_0 \mathcal{A}^\tau(T, z)) \leq C_T,$$

and, moreover, for every τ and for every z with $\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$, the affine-Volterra Riccati system (3.3) and (3.4) is well posed on $[0, T]$ and $z \mapsto \mathcal{B}^\tau(T, z) + V_0 \mathcal{A}^\tau(T, z)$ is analytic on the open strip $\{\frac{1}{2} - \delta < \operatorname{Re}(z) < \frac{1}{2} + \delta\}$.

Proof. We provide the proof in Appendix B.3. □

Corollary 3.6 (Moment strip). *Under the assumptions in Lemma 3.3 and Proposition 3.5, the following holds. There exists $\delta > 0$ such that for all z with $\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$ the Riccati systems (3.1) and (3.2) are well-posed on $[0, T]$ and*

$$\sup_{\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \operatorname{Re}(\mathcal{B}(T, z) + V_0 \mathcal{A}(T, z)) < \infty,$$

and the same bound holds for $(\mathcal{A}^\tau, \mathcal{B}^\tau)$ uniformly in τ .

Remark 3.7 (Practical validity of the strip and (3.8)). We verify that for the parameters used in our experiments (Table 2), the conditions for Lemma 3.3 and Proposition 3.5 are satisfied.

1. (Lemma 3.3): We check the strip width δ . Recall under the canonical rough Heston parametrization $c_X = 1, c_V = \nu$. With the parameters from Table 2: $\alpha = 0.62, \gamma = 0.1, \nu = 0.331, T = 1$, we can evaluate the kernel integrals as:

$$G(1) = \int_0^1 (f_{\alpha,\gamma}(s))^2 ds \approx 0.8787, \quad F(1) = \int_0^1 f_{\alpha,\gamma}(s) ds \approx 0.7147.$$

Since $f_{\alpha,\gamma} \geq 0$, the L^∞ -norm is achieved at $s = 0$: $\|\Phi\|_{L^\infty(0,T)} = F(T)$. First, we compute the threshold Θ_T :

$$\Theta_T = \frac{2}{c_V^2 \|\Phi\|_{L^\infty(0,T)}^2} = \frac{2}{\nu^2 F(T)^2} \approx \frac{2}{(0.331)^2 (0.7147)^2} \approx \frac{2}{0.05596} \approx 35.74.$$

Next, we compute the key term for the strip width:

$$\frac{2\Theta_T}{c_X^2} = \frac{2\Theta_T}{1^2} = 2\Theta_T \approx 2 \times 35.74 \approx 71.48.$$

Using the formula from Lemma 3.3, the half-width δ (centered at $y = 1/2$) is:

$$\delta = \frac{\sqrt{1 + 2\Theta_T/c_X^2} - 1}{4} \approx \frac{\sqrt{1 + 71.48} - 1}{4} = \frac{\sqrt{72.48} - 1}{4} \approx \frac{8.51 - 1}{4} \approx 1.88.$$

Since $\delta \approx 1.88$, a strip of half-width $\delta = 1/2$ (which is sufficient for the Carr-Madan formula and implies $\mathbb{E}[S_T] < \infty$) exists and is guaranteed by a very wide margin.

2. (Condition (3.8)): We still need the uniform L^2 -energy bound for Proposition 3.5. Since $\zeta^\tau \rightarrow f_{\alpha,\gamma}$ in $L^2(0, T)$ (by Proposition 3.8), the sequence $\int_0^T (\zeta^\tau(s))^2 ds$ is convergent and thus bounded. Therefore, $\sup_{\tau \geq 1} \int_0^T (\zeta^\tau(s))^2 ds < \infty$ holds, and the proposition applies.

We now quantify the error of the step kernel approximation on a finite horizon, by deriving an explicit L^2 rate (and an integrated L^1 bound) from Lemma 3.1.

Proposition 3.8 (Finite-horizon L^2 kernel rate (step approximation)). *Fix $\alpha \in (\frac{1}{2}, 1)$ and $\gamma > 0$, and assume $1 - a_\tau = \gamma\tau^{-\alpha}$. Under Lemma 3.1, i.e. $\hat{\varphi}(s) = 1 - s^\alpha + R(s)$ with $|R(s)| \leq C_R |s|^{\alpha+\delta}$ for some $\delta \in (0, 1 - \alpha]$, we have*

$$\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,1)} \leq C_1 \tau^{\frac{1}{2}-\alpha},$$

for a constant C_1 depending only on $(\alpha, \gamma, \delta, C_R)$. Moreover,

$$\sup_{t \in [0,1]} \int_0^t |\zeta^\tau(u) - f_{\alpha,\gamma}(u)| du \leq C_2 \tau^{\frac{1}{2}-\alpha},$$

for a constant C_2 depending only on $(\alpha, \gamma, \delta, C_R)$. (Equivalently, the same bounds hold on $[0, T]$ for any $T \in (0, 1]$, with constants uniform in $T \leq 1$.)

Proof. We provide the proof in Appendix B.4. □

We next show that the microscopic drivers admit a Brownian coupling with time-averaged L^2 accuracy, relying only on their martingale representation and bracket identity.

Proposition 3.9 (Microscopic invariance, time-averaged L^2). *For the microscopic INAR/compound-Poisson layers with representation and bracket:*

$$M_t^\tau = \tau^{-1/2} \sum_{k \leq \lfloor t\tau \rfloor} \xi_k^\tau, \quad [M^\tau]_t = \tau^{-1} \sum_{k \leq \lfloor t\tau \rfloor} v_k^\tau, \quad (3.9)$$

where $(\xi_k^\tau)_{k \geq 1}$ is an \mathcal{F}_k^τ -martingale difference array (i.e. $\mathbb{E}[\xi_k^\tau | \mathcal{F}_{k-1}^\tau] = 0$) and $v_k^\tau := \mathbb{E}[(\xi_k^\tau)^2 | \mathcal{F}_{k-1}^\tau]$, there exists a Brownian motion W on a common space such that, for each fixed $T > 0$,

$$\int_0^T \mathbb{E}[|M_t^\tau - W_t|^2] dt \leq C_T \tau^{-1/2}, \quad \sup_{t \in [0, T]} |\mathbb{E}[M^\tau]_t - t| \leq C_T \tau^{-1}.$$

Proof. We provide the proof in Appendix B.5. □

We next transfer the microscopic Brownian coupling to Fourier-damped price components and obtain a $\tau^{1/4}$ rate. For any $u \in \mathbb{R}$, set $z(u) := \frac{1}{2} + iu \in \mathbb{C}$. For a terminal log-price X_T and a log-strike $k \in \mathbb{R}$, define the Fourier-damped price component by

$$\text{Price}_{u,k}(X_T) := \mathbb{E}\left[e^{z(u)(X_T - k)}\right] = e^{-z(u)k} \phi_{X,T}(z(u)), \quad (3.10)$$

where $\phi_{X,T}(z) := \mathbb{E}[e^{zX_T}]$, whenever $z(u)$ lies in the risk-neutral moment strip.

Theorem 3.10 (Microscopic \rightarrow SDE reduction with $\tau^{1/4}$ -rate in price). *Let (S^τ, V^τ) denote the ζ^τ -Volterra stochastic volatility SDE, i.e., the analogue of the limiting system (3.5)-(3.6) but with the kernel $f_{\alpha,\gamma}$ replaced by ζ^τ , driven by a Brownian motion $W = (W^X, W^V)$ defined in (2.8)-(2.9). Let $(\hat{S}^\tau, \hat{V}^\tau)$ be the same SDE (same coefficients and initial data) but driven by the microscopic martingale $M^\tau = (M^{\tau,X}, M^{\tau,V})$ on the same space such that, by Proposition 3.9, the microscopic invariance estimate holds*

$$\int_0^T \mathbb{E}[|M_t^\tau - W_t|^2] dt \leq C_T \tau^{-1/2}. \quad (3.11)$$

Then, for any compact intervals $U \subset \mathbb{R}$ and $K \subset \mathbb{R}$ such that $z(u) = \frac{1}{2} + iu$ lies in the moment strip for all $u \in U$, there exists a constant $C_{T,U,K} < \infty$ with

$$\sup_{u \in U, k \in K} \left| \text{Price}_{u,k}(\log \hat{S}_T^\tau) - \text{Price}_{u,k}(\log S_T^\tau) \right| \leq C_{T,U,K} \tau^{-1/4}.$$

In particular, when $\alpha \in (\frac{1}{2}, \frac{3}{4})$ (so that $\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)} \lesssim \tau^{\frac{1}{2}-\alpha}$), the microscopic error $\mathcal{O}(\tau^{-1/4})$ is negligible compared to the kernel error $\mathcal{O}(\tau^{\frac{1}{2}-\alpha})$; when $\alpha \in [\frac{3}{4}, 1)$ the microscopic error dominates and the overall rate is $\mathcal{O}(\tau^{-1/4})$.

Proof. We provide the proof in Appendix B.6. \square

3.2 European options

Fix $z \in \mathbb{C}$ with $\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$. We introduce the auxiliary function $H(\cdot; z)$ and re-write the fractional Riccati–Volterra system (3.1) and (3.2) as

$$H(t; z) = \int_0^t f_{\alpha,\gamma}(t-s) \left(\frac{1}{2}(z^2 - z) + (\rho\xi z - \kappa) H(s; z) + \frac{1}{2}\xi^2 (H(s; z))^2 \right) ds, \quad t \in [0, T]. \quad (3.12)$$

Then

$$\mathcal{A}(t; z) = (f_{\alpha,\gamma} * H(\cdot; z))(t), \quad \mathcal{B}(t; z) = \kappa\theta \int_0^t H(s; z) ds, \quad (3.13)$$

and the same formulas hold with $f_{\alpha,\gamma}$ replaced by ζ^τ for $(\mathcal{A}^\tau, \mathcal{B}^\tau)$.

To propagate kernel approximation errors to characteristic exponents, we establish a Lipschitz stability estimate for the fractional Riccati mapping.

Proposition 3.11 (Lipschitz stability of the fractional Riccati mapping). *Under Corollary 3.6 and Proposition 3.8, there exists $C_T < \infty$ (locally uniform in z on the strip $\{z \in \mathbb{C} : \operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]\}$ guaranteed by Corollary 3.6) such that for all $T > 0$,*

$$\begin{aligned} & \sup_{t \in [0, T]} (|\mathcal{A}^\tau(t; z) - \mathcal{A}(t; z)| + |\mathcal{B}^\tau(t; z) - \mathcal{B}(t; z)|) \\ & \leq C_T (\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T)} + \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)} + \|F^\tau - F\|_{L^\infty(0,T)}). \end{aligned}$$

In particular, by Proposition 3.8, $\sup_{t \leq T} (|\mathcal{A}^\tau - \mathcal{A}| + |\mathcal{B}^\tau - \mathcal{B}|) = \mathcal{O}(\tau^{\frac{1}{2}-\alpha})$.

Proof. We provide the proof in Appendix B.7. \square

We next quantify how perturbations of the kernel (and its primitive) propagate to the Riccati–Volterra solution, yielding Lipschitz control uniformly on the moment strip.

Proposition 3.12 (From Riccati stability to CF stability). *Under the assumptions in Corollary 3.6, for all z in the strip $\{z \in \mathbb{C} : \operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]\}$, the following holds:*

$$|\phi_T^\tau(z) - \phi_T(z)| \leq C_T(z) (|\mathcal{B}^\tau(T, z) - \mathcal{B}(T, z)| + V_0 |\mathcal{A}^\tau(T, z) - \mathcal{A}(T, z)|),$$

with $C_T(z)$ locally bounded on $\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$.

Proof. We provide the proof in Appendix B.8. \square

We express price errors via the Carr–Madan Fourier map (see Carr and Madan (1999)) on the damping line $\operatorname{Re}(z) = \frac{1}{2}$.

Proposition 3.13 (Carr–Madan map). *Let $K > 0$, $P_T := \log(S_T/K)$ and $\phi_{P,T}(z) := \mathbb{E}[e^{zP_T}]$. Then*

$$|C^{\text{aux},\tau}(T, K) - C(T, K)| \leq \frac{K}{2\pi} \int_{\mathbb{R}} \frac{|\phi_{P,T}^\tau(\frac{1}{2} + iu) - \phi_{P,T}(\frac{1}{2} + iu)|}{\frac{1}{4} + u^2} du.$$

Equivalently, if one works with $X_T := \log S_T$ and $\phi_{X,T}(z) := \mathbb{E}[e^{zX_T}]$, then using $\phi_{P,T}(z) = K^{-z} \phi_{X,T}(z)$,

$$|C^{\text{aux},\tau}(T, K) - C(T, K)| \leq \frac{\sqrt{K}}{2\pi} \int_{\mathbb{R}} \frac{|\phi_{X,T}^\tau(\frac{1}{2} + iu) - \phi_{X,T}(\frac{1}{2} + iu)|}{\frac{1}{4} + u^2} du.$$

Proof. We provide the proof in Appendix B.9. \square

Now, we are ready to present the weak error bound for European option pricing.

Theorem 3.14 (Weak error bound for European call). *Under the assumptions in Corollary 3.6 and Proposition 3.8, for each fixed $T \in (0, 1]$ and $K > 0$, the total weak error satisfies*

$$|C^{\text{INAR},\tau}(T, K) - C(T, K)| \leq C_{T,K} \left(\tau^{\frac{1}{2}-\alpha} + \tau^{-1/4} \right),$$

where the constant $C_{T,K}$ is locally uniform in K and in model parameters on compact sets contained in the moment strip. In particular,

$$|C^{\text{INAR},\tau}(T, K) - C(T, K)| = \begin{cases} \mathcal{O} \left(\tau^{\frac{1}{2}-\alpha} \right), & \alpha \in (\frac{1}{2}, \frac{3}{4}), \\ \mathcal{O} \left(\tau^{-1/4} \right), & \alpha \in [\frac{3}{4}, 1). \end{cases}$$

Proof. We provide the proof in Appendix B.10. \square

3.3 Arithmetic Asian options

Let

$$\bar{S}_T := \frac{1}{T} \int_0^T S_t dt, \quad AA(T, K) := \mathbb{E} \left[(\bar{S}_T - K)^+ \right],$$

and define analogously \bar{S}_T^τ and $AA^{\text{aux},\tau}(T, K)$ for the auxiliary ζ^τ -Volterra model, and $AA^{\text{INAR},\tau}(T, K)$ for the microscopic version driven by M^τ (same coefficients and kernel ζ^τ).

We first record a simple Lipschitz property which reduces the weak error for arithmetic Asians to a time-averaged path difference.

Lemma 3.15 (Time-averaging Lipschitz property). *The following inequality holds:*

$$|AA^{\text{aux},\tau}(T, K) - AA(T, K)| \leq \frac{1}{T} \int_0^T \mathbb{E} |S_t^\tau - S_t| dt.$$

Proof. We provide the proof in Appendix B.11. \square

To control tails needed for Fourier methods, we now quantify how far exponential moments extend beyond order 1 under a Novikov-type bound.

Lemma 3.16 (Exponential moments beyond order 1). *Fix $T > 0$. Let X be the log-price of the affine-Volterra (rough Heston-type) model on $[0, T]$ with diffusion loadings $c_X > 0$ in the X -equation (3.5) and $c_V > 0$ in the V -equation (3.6), and Volterra kernel $f_{\alpha,\gamma}$. Set*

$$\Phi(s) := \int_0^{T-s} f_{\alpha,\gamma}(u) du, \quad \Theta_T := \frac{2}{c_V^2 \|\Phi\|_{L^\infty(0,T)}^2}.$$

Define the effective threshold $\Theta_{\text{eff}} := \Theta_T/4$. Then:

(a) (Limit model). *For every $y \in \mathbb{R}$ with $y \geq 0$ satisfying*

$$c_X^2 (2y^2 - y) < \Theta_{\text{eff}}, \tag{3.14}$$

one has

$$\sup_{t \in [0, T]} \mathbb{E} [e^{yX_t}] < \infty. \tag{3.15}$$

Equivalently, the set of admissible y is the interval $0 \leq y < y_{\max}$, where

$$y_{\max} := \frac{1 + \sqrt{1 + 8\Theta_{\text{eff}}/c_X^2}}{4} = \frac{1 + \sqrt{1 + 2\Theta_T/c_X^2}}{4},$$

so that, in particular, $\mathbb{E}[e^{(1+\theta)X_t}] < \infty$ for $t \in [0, T]$ and $\theta \in (0, \theta^*)$, with

$$\theta^* := y_{\max} - 1 = \frac{\sqrt{1 + 2\Theta_T/c_X^2} - 3}{4}.$$

This requires $\sqrt{1 + 2\Theta_T/c_X^2} > 3$, or $\Theta_T > 4c_X^2$, to have $\theta^* > 0$.

(b) (Uniform-in- τ prelimit). Let ζ^τ be a family of causal approximating kernels with cumulative $F^\tau(t) = \int_0^t \zeta^\tau(s)ds$ and write $\Phi^\tau(s) := F^\tau(T-s)$. Suppose the uniform energy bound

$$\sup_{\tau \geq 1} \int_0^T (\zeta^\tau(s))^2 ds < \infty,$$

so that $\sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0,T)} < \infty$, and define

$$\Theta^* := \frac{2}{c_V^2 \sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0,T)}^2} > 0, \quad \Theta_{\text{eff}}^* := \Theta^*/4.$$

Then for every $y \geq 0$ with

$$c_X^2 (2y^2 - y) < \Theta_{\text{eff}}^*, \tag{3.16}$$

one has

$$\sup_{\tau \geq 1} \sup_{t \in [0, T]} \mathbb{E} \left[e^{yX_t^{(\tau)}} \right] < \infty.$$

(c) (Coarse sufficient conditions). Using $\|\Phi\|_{L^\infty(0,T)} \leq \sqrt{T} \left(\int_0^T (f_{\alpha, \gamma}(s))^2 ds \right)^{1/2}$, a sufficient condition for (3.14) is

$$c_X^2 c_V^2 T \left(\int_0^T (f_{\alpha, \gamma}(s))^2 ds \right) (2y^2 - y) < \frac{1}{2}. \tag{3.17}$$

Likewise, a sufficient condition for (3.16) is

$$c_X^2 c_V^2 T \left(\sup_{\tau \geq 1} \int_0^T (\zeta^\tau(s))^2 ds \right) (2y^2 - y) < \frac{1}{2}.$$

In particular, in the canonical rough Heston case $c_X = 1, c_V = \nu$ this reduces to

$$\nu^2 T \left(\int_0^T (f_{\alpha, \gamma}(s))^2 ds \right) (2y^2 - y) < \frac{1}{2} \quad (\text{limit model}),$$

and similarly for the uniform prelimit.

Proof. We provide the proof in Appendix B.12. □

Remark 3.17 (Practical validity of exponential moment conditions). The condition $\theta^* > 0$ (i.e., that the moment strip extends beyond $y = 1$) simply requires $\Theta_T > 4c_X^2$. As demonstrated in our numerical validation (Remark 3.7), for typical rough Heston parameters ($\nu = 0.331, T = 1$), we have $\Theta_T \approx 35.74$ while $4c_X^2 = 4 \cdot 1^2 = 4$. The condition $\Theta_T > 4c_X^2$ is satisfied by a very large margin.

Furthermore, our weak error bound for Asian options (Theorem 3.18) requires the existence of 2-nd order exponential moments, i.e., $\mathbb{E}[e^{2X_T}] < \infty$. Using the condition from Lemma 3.16(a), this requires:

$$c_X^2 (2y^2 - y) < \Theta_{\text{eff}}, \quad \text{at } y = 2,$$

where $\Theta_{\text{eff}} = \Theta_T/4$. Plugging in $y = 2$, the condition becomes:

$$c_X^2 (2(2^2) - 2) < \Theta_T/4 \implies 6c_X^2 < \Theta_T/4 \implies 24c_X^2 < \Theta_T.$$

Using our numerical parameters again, we compute the LHS and RHS of the condition (3.14) (evaluated at $y = 2$):

$$\text{LHS} = 24c_X^2 = 24 \cdot 1^2 = 24, \quad \text{RHS} = \Theta_T \approx 35.74.$$

Since $24 < 35.74$, the condition for 2-nd order moments is also satisfied by a wide margin. This confirms that the exponential moment assumptions in our theorems are non-restrictive for practical applications.

Now, we are ready to present the weak error bound for arithmetic Asian option pricing.

Theorem 3.18 (Weak error bound for arithmetic Asian call). *Under the assumptions of Proposition 3.8, and assuming the order-2 exponential-moment condition of Lemma 3.16 holds uniformly in τ , i.e.*

$$\sup_{t \in [0, T]} \mathbb{E} [e^{2X_t}] < \infty \quad \text{and} \quad \sup_{\tau \geq 1} \sup_{t \in [0, T]} \mathbb{E} [e^{2X_t^\tau}] < \infty. \quad (3.18)$$

Then for each fixed $T \in (0, 1]$ and $K > 0$,

$$|AA^{\text{INAR}, \tau}(T, K) - AA(T, K)| \leq C_{T, K} \left(\tau^{\frac{1}{2} - \alpha} + \tau^{-1/4} \right),$$

where the constant $C_{T, K}$ is locally uniform in K and in model parameters on compact sets satisfying the stated exponential-moment condition.

Proof. We provide the proof in Appendix B.13. □

Remark 3.19. Using Lemma 3.16 with $y = 2$, a sufficient condition for (3.18) to hold is to assume $24c_X^2 < \Theta_T$ for the limit model and $24c_X^2 < \Theta^*$ for the prelimit, where $\Theta^* := \frac{2}{c_V^2 \sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0, T)}^2}$ (coarse sufficient conditions from Lemma 3.16(c) with $y = 2$: $c_X^2 c_V^2 T G(T) < 1/12$ and $c_X^2 c_V^2 T \sup_{\tau \geq 1} G^\tau(T) < 1/12$).

3.4 Lookback options

We consider the fixed-strike lookback call option with payoff

$$LB(T, K) := \mathbb{E} \left[\left(\sup_{0 \leq t \leq T} S_t - K \right)^+ \right],$$

and define analogously $LB^{\text{aux}, \tau}(T, K)$ for the auxiliary ζ^τ -Volterra model and $LB^{\text{INAR}, \tau}(T, K)$ for the microscopic version driven by M^τ (same coefficients and kernel ζ^τ). Recall $S_t = e^{X_t}$ and note that $\sup_{t \leq T} S_t = e^{\sup_{t \leq T} X_t}$.

The next statement gives the *auxiliary SDE* weak error for lookbacks on a finite horizon, under the standing assumptions of Section 3 (moment strip around $\text{Re}(z) = \frac{1}{2}$, well-posed Riccati/CF, and the L^2 kernel rate), without any additional exponential-moment or boundary/bridge corrections.

Proposition 3.20 (Auxiliary SDE rate for lookbacks). *Under the assumptions in Lemma 3.16, Corollary 3.6 and Proposition 3.8, for each fixed $T \in (0, 1]$ and $K > 0$,*

$$|LB^{\text{aux}, \tau}(T, K) - LB(T, K)| \leq C_{T, K} \tau^{\frac{1}{4} - \frac{\alpha}{2}}, \quad \tau \rightarrow \infty,$$

where the constant $C_{T, K}$ is locally uniform on compact parameter sets contained in the moment strip.

Proof. We provide the proof in Appendix B.14. □

The microscopic layer contributes additively through the martingale driver coupling. Combining Proposition 3.20 with the microscopic reduction (Theorem 3.10) yields the following total weak error bound for lookback option pricing.

Theorem 3.21 (Weak error bound for lookback call). *Fix $T \in (0, 1]$ and $K > 0$, and let $\alpha \in (\frac{1}{2}, 1)$, $\gamma > 0$. Assume:*

- (a) *the exponential-moment condition in Lemma 3.16 holds (and in particular the Novikov-type threshold of Lemma 3.3 is satisfied);*
- (b) *the kernel approximation satisfies the step scheme $1 - a_\tau = \gamma\tau^{-\alpha}$ and the finite-horizon L^2 rate of Proposition 3.8.*

Then

$$|LB^{\text{INAR}, \tau}(T, K) - LB(T, K)| \leq C_{T, K} \left(\tau^{\frac{1}{4} - \frac{\alpha}{2}} + \tau^{-1/8} \sqrt{\log \tau} \right),$$

where the constant $C_{T, K}$ is locally uniform on compact parameter sets contained in the moment strip and is independent of τ . In particular, for $\alpha \in (\frac{1}{2}, \frac{3}{4})$ the kernel term dominates, whereas for $\alpha \in [\frac{3}{4}, 1)$ the microscopic contribution $\tau^{-1/8} \sqrt{\log \tau}$ dominates.

Proof. We provide the proof in Appendix B.15. □

Remark 3.22 (Exponential moment condition). The proof of Theorem 3.21 relies on Proposition 3.20, which uses Doob's L^p inequality to control the running maximum. This step requires an exponential moment of order $p > 1$, specifically $\sup_{t \leq T} \mathbb{E}[e^{pX_t}] < \infty$ (and the corresponding bound for the prelimit models).

This remark clarifies why the assumption in item (a) of Theorem 3.21—namely the exponential-moment condition of Lemma 3.16, which is derived from Lemma 3.3—is sufficient, and why this assumption differs from the one made for Asian options (Theorem 3.18).

1. Lemma 3.3 (the Novikov-type threshold) yields the quantitative bound Θ_T defined in (3.7).
2. As shown in Lemma 3.16(a), this Θ_T determines the maximal available exponential moment $y_{\max} = \frac{1 + \sqrt{1 + 2\Theta_T/c_X^2}}{4}$.
3. As demonstrated in Remark 3.7 for our parameter choices, this Θ_T is large enough to ensure $y_{\max} > 1$.
4. Therefore, the conclusion of Lemma 3.16 (assumption (a) in the theorem) guarantees the existence of some $p \in (1, y_{\max})$ such that $\sup_{t \leq T} \mathbb{E}[e^{pX_t}] < \infty$, which is exactly what is needed to apply Doob's L^p inequality in the lookback option proof.
5. This contrasts with the Asian option case (Theorem 3.18), where the proof requires a *specific* moment of order $p = 2$. Since Lemma 3.3 only guarantees $y_{\max} > 1$ (but not necessarily $y_{\max} \geq 2$), a stronger, explicit assumption of order-2 exponential moments was imposed there.

4 Numerical Results

Pricing framework and risk-neutral measure. We work under the risk-neutral measure \mathbb{Q} associated with the money-market numéraire $B_t^0 = \exp\left(\int_0^t r_u du\right)$ where $(r_t)_{t \in [0, 1]}$ represents the interest rate process. Let the discounted price be $\tilde{S}_t := S_t/B_t^0$. For notational convenience, we keep writing S_t for \tilde{S}_t . Under this convention, S is a \mathbb{Q} -martingale and evolves as

$$dS_t = S_t \sqrt{V_t} dW_t^X,$$

while V_t follows the rough Heston dynamics specified below, and $d\langle W^X, W^V \rangle_t = \rho dt$. Equivalently, one may view S_t as a (discounted or T -forward) price process with zero drift under the corresponding pricing measure.

This implies that the discrete-time INAR(∞) process, for a sufficiently large τ , serves as a numerical approximation to the risk-neutral dynamics. All option prices are therefore computed by simulating these approximating paths and calculating the discounted expected payoff under this implicitly defined risk-neutral measure. For simplicity and consistency with the cited literature, we assume the risk-free interest rate $r = 0$; hence no discounting is required in our primary results. Our approach is to posit the limiting model directly under \mathbb{Q} and use the INAR(∞) process as an effective simulation scheme for it.

Based on the theoretical convergence established above, we can simulate a rough Heston model by simulating from its approximate cumulative INAR(∞) model. More specifically, we present an efficient Monte Carlo simulation framework for pricing various types of path-independent and path-dependent options by simulating a two-dimensional cumulative INAR(∞) processes and demonstrate numerically that they provide accurate approximations to corresponding option prices under the rough Heston model. Our implementation includes European, Asian, lookback, and barrier options, with an emphasis on computational efficiency through parallel processing.

We simulate the INAR(∞) recursion on $[0, T]$ with zero pre-history ($X_n \equiv 0$ for $n \leq 0$) via a non-circular linear convolution (CDQ-FFT). At time $n \leq \lfloor \tau T \rfloor$, the update only involves lags $1, \dots, n-1$. Hence, for any $p \geq n-1$ (in particular $p \geq \lfloor \tau T \rfloor$), the INAR(p) recursion coincides *pathwise* with our INAR(∞) update on $[0, T]$. In other words, our implementation does not fix p ; effectively it uses all available lags at each step, and therefore no extra “ $p \rightarrow \infty$ truncation” error arises on the finite horizon.

Our theoretical framework is based on the limit as the discretization parameter $\tau \rightarrow \infty$. In any practical implementation, a finite value for τ must be chosen. This parameter represents the number of discrete time steps in our simulation over the unit time horizon ($T = 1$) and governs the accuracy of the approximation to the continuous-time limit.

The choice of τ involves a trade-off between accuracy and computational cost. A naive implementation of the INAR(∞) simulation has a per-path complexity of $\mathcal{O}(\tau^2)$. To improve this, we implement a **divide-and-conquer (CDQ) FFT scheme**; see Algorithm 1. The recursion first simulates the left half of the interval, applies a single FFT-based convolution to propagate its contribution to the right half, and then repeats on the right. The work per level is $\mathcal{O}(n \log n)$ for interval length n , so that the recurrence $T(n) = 2T(n/2) + \mathcal{O}(n \log n)$ yields an overall per-path complexity $T(\tau) = \mathcal{O}(\tau \log^2 \tau)$. In practice, the constants remain small thanks to FFT reuse and thread-level parallelism, giving substantial speedups over the naive $\mathcal{O}(\tau^2)$ loop while preserving numerical stability.

4.1 Rough case ($\alpha = 0.62$)

As established in the preamble to this section, we work under a risk-neutral measure \mathbb{Q} , under which the asset price S_t and its variance process V_t follow the rough Heston model dynamics.

We consider the following four types of options, each defined by its specific payoff function V_T :

1. **European Option:** The payoff depends only on the asset price at maturity, S_T :

$$H_T = \begin{cases} \max(S_T - K, 0) & \text{for a call option,} \\ \max(K - S_T, 0) & \text{for a put option.} \end{cases}$$

2. **Arithmetic Asian Option:** The payoff depends on the arithmetic average of the asset price over the life of the option:

$$H_T = \begin{cases} \max(\bar{S}_T - K, 0) & \text{for a call option,} \\ \max(\bar{K} - S_T, 0) & \text{for a put option,} \end{cases}$$

where

$$\bar{S}_T = \frac{1}{M+1} \sum_{i=0}^M S_{i\Delta t}, \quad \text{with } \Delta t = T/M. \quad (4.1)$$

In the literature, one often studies the continuous approximation $\frac{1}{T} \int_0^T S_u du$ of \bar{S}_T for mathematical convenience. In simulation, we work with the time average (4.1) with the grid values including the initial point directly.

3. **Lookback Option:** The payoff depends on the maximum ($M_T := \max_{0 \leq t \leq T} S_t$) or minimum ($m_T := \min_{0 \leq t \leq T} S_t$) price achieved during the option's life:

$$H_T = \begin{cases} M_T - K & \text{for a lookback call option,} \\ K - m_T & \text{for a lookback put option.} \end{cases}$$

4. **Barrier Options:** The payoff is contingent on whether the asset price has reached a certain barrier level B during the option's life. For an up-and-in barrier call option, the payoff is:

$$H_T = \begin{cases} \max(S_T - K, 0) & \text{if } \max_{0 \leq t \leq T} S_t \geq B, \\ 0 & \text{otherwise.} \end{cases}$$

For a down-and-out barrier put option, the payoff is:

$$H_T = \begin{cases} \max(K - S_T, 0) & \text{if } \min_{0 \leq t \leq T} S_t > B, \\ 0 & \text{otherwise.} \end{cases}$$

The core of our implementation consists of several key functions that work together to simulate price paths and compute option values. The primary components include the following.

1. A helper function $f_\alpha(t)$ that computes the fractional kernel:

$$f_\alpha(t) = \begin{cases} 0 & \text{if } t \leq 0, \\ 1 - \frac{1}{\Gamma(1-\alpha)} & \text{if } t = 1, \\ \frac{1}{\Gamma(1-\alpha)} \left(\frac{1}{t^\alpha} - \frac{1}{(t+1)^\alpha} \right) & \text{if } t > 1. \end{cases}$$

2. Path generation functions for different option types:

- Standard path simulation for European options;
- Path with running average calculation for Asian options;
- Path with maximum/minimum tracking for lookback options;
- Barrier crossing detection for barrier options.

4.1.1 Enhanced parallel Monte Carlo implementation

We employ a parallel Monte Carlo simulation that effectively distributes the computational workload across multiple CPU cores. The structure of the algorithm is described in Algorithm 1.

To empirically demonstrate the convergence of our scheme, we performed a series of tests for a European call option with strike $K = 100$. Table 3 shows the estimated option price, its 95% confidence interval, and the corresponding CPU time for a range of increasing τ values.

Table 3 indicates statistical consistency of our scheme: as τ increases from 40 to 320, the estimates lie within narrow 95% confidence intervals around the benchmark and stabilize. Point estimates are monotone in τ .

Algorithm 1: Parallel INAR(∞) Simulation with CDQ FFT Convolution

Input: Model Parameters $(\alpha, \gamma, \rho, \nu, V_0, S_0)$, Simulation Parameters (τ, n_{sims})
Output: Price estimates and 95% CIs for European, Asian, Lookback, and Barrier options
Preprocessing:
Derive parameters $\beta \leftarrow \text{solve_beta}(\rho)$, $\mu \leftarrow \text{solve_mu}(\nu, \theta, \beta, \gamma)$, $\xi \leftarrow \text{solve_xi}(V_0)$ Derive INAR(∞) time-scaled parameters a_τ, μ_τ from $(\gamma, \mu, \tau, \alpha)$ Pre-compute the fractional kernel $w_k = \frac{a_\tau}{1+\beta} f_\alpha(k)$ for $k = 1, \dots, \tau$ Pre-compute the baseline intensity $\hat{\mu}_\tau(t)$ for $t = 1, \dots, \tau$
Parallel Simulation:
Determine number of available CPU threads N_{threads} Partition n_{sims} simulations among threads **for** each thread in parallel **do**
 Initialize thread-specific random number generator **for** each simulation assigned to thread **do**
 Initialize arrays X^\pm, N^\pm, S , history vector λ^{hist} , and data vector Y // CDQ recursion on $[1, \tau]$
 Function CDQ((L, R))
 if $L = R$ **then**
 $\lambda \leftarrow \max\{0, \hat{\mu}_\tau(L) + \lambda_L^{\text{hist}}\}$ Sample $X_L^+, X_L^- \sim \text{Poisson}(\lambda)$ independently Update N_L^\pm , price S_L , and set $Y_L = X_L^+ + \beta X_L^-$ **return**
 end
 $M \leftarrow \lfloor (L + R)/2 \rfloor$ // 1. Simulate left half
 CDQ(L, M) // 2. Convolution from left to right via FFT
 Form vectors $A = Y_L, \dots, Y_M$ and $B = w_1, \dots, w_{R-L}$ Compute $\text{IFFT}(\text{FFT}(A) \cdot \text{FFT}(B))$ to obtain contribution Δ Add Δ to $\lambda_{M+1}^{\text{hist}}, \dots, \lambda_R^{\text{hist}}$ // 3. Recurse on right half
 CDQ($M+1, R$)
 end
 CDQ($1, \tau$) Compute payoffs for all option types from the full simulated path S Store the payoffs
 end
end
Postprocessing:
Aggregate all stored payoffs from all threads For each option type, compute the mean and 95% confidence interval from its payoffs **return** A structure containing all option price estimates and their CIs

Balancing accuracy and cost, we therefore use $\tau = 320$ as a practical default: further increases yield improvements that are within sampling error while the computational time grows noticeably. We keep $\tau = 320$ for the remaining experiments unless otherwise stated.

We conducted extensive numerical experiments using the parameter settings shown in Table 2, which are the same as Callegaro et al. (2021).

Parameter	Value	Type	Derivation
α	0.62	Direct	-
γ	0.1	Direct	-
ρ	-0.681	Direct	-
ν	0.331	Direct	vol-of-vol in v -equation
θ	0.3156	Direct	-
V_0	0.0392	Direct	-
S_0	100	Direct	-
β	27.5583	Derived	From $\rho = \frac{1 - \beta}{\sqrt{2(1 + \beta^2)}}$
μ	26.8592	Derived	From $\mu = \frac{\theta(1 + \beta^2)}{\gamma\nu^2(1 + \beta)^2}$
ξ	0.124208	Derived	From $V_0 = \xi\theta$

Table 2: Model parameters for numerical experiments.

4.1.2 Convergence Analysis and Optimal Discretization

Our theoretical framework is based on the convergence of the discrete INAR(∞) process to the continuous-time rough Heston model as the number of time steps, τ , approaches infinity. In any practical implementation, a finite value for τ must be chosen. This parameter represents the granularity of our discretization and critically governs the accuracy of the approximation.

The choice of τ involves a non-trivial trade-off between two opposing sources of error: the *discretization error* and a *model specification bias* inherent in the finite- τ parameterization. To empirically investigate this trade-off, we conducted a series of convergence tests for a European call option with strike $K = 100$, using an increasing number of time steps from $\tau = 40$ to $\tau = 320$.

Table 3: Convergence of the European call option price ($K = 100$) with respect to the discretization parameter τ . All simulations were run with 500,000 paths.

Parameter τ	Estimated Price	Deviation from Benchmark	Computation Time(Seconds)
40	9.5160	+0.0423	5.5774s
80	9.4473	-0.0264	12.5218s
160	9.4540	-0.0197	27.8627s
320	9.4898	-0.0161	70.3283s
Benchmark (Callegaro et al. (2021))		9.4737	

The following Table 4 presents the numerical results for all option types.

For the case of European options pricing, the values in parentheses (22.1366, 14.9672, 9.4737, 5.6234, and 3.1424) in Table 4 represent the results from Callegaro et al. (2021). These values appear in the Call column under European Options and serve as benchmark values for comparison with our current research results. Our study used the same parameter settings as Callegaro et al. (2021) (as shown in Table 2), enabling direct comparison of the results. As evident from Table 4, our results closely match those of Callegaro et al. (2021), demonstrating the accuracy of our methodology. The small differences between our results and the benchmark values (typically within a few basis points) validate the effectiveness of our INAR(∞) approach for option pricing under the rough Heston model. Note that the method of Callegaro et al. (2021) is analytical in nature and is based on solving the fractional Riccati equation, and they are used for benchmarking purposes. Note that we have not been able to find benchmark numerical results for the valuation of Asian, lookback or barrier options under the rough Heston model.

On a Mac mini M4, our implementation completes in about 70.3283 seconds leveraging parallel computing

European Options				
Strike	Call	95% CI	Put	95% CI
80	22.1558(22.1366)	[22.0972, 22.2145]	2.1141	[2.0989, 2.1294]
90	14.9839(14.9672)	[14.9324, 15.0355]	4.9422	[4.9177, 4.9668]
100	9.4898(9.4737)	[9.4470, 9.5326]	9.4481	[9.4136, 9.4825]
110	5.6407(5.6234)	[5.6070, 5.6744]	15.5990	[15.5553, 15.6427]
120	3.1584(3.1424)	[3.1331, 3.1837]	23.1167	[23.0654, 23.1680]

Asian Options				
Strike	Call	95% CI	Put	95% CI
80	20.2329	[20.1980, 20.2678]	0.2143	[0.2108, 0.2177]
90	11.4845	[11.4540, 11.5151]	1.4658	[1.4557, 1.4760]
100	5.1763	[5.1539, 5.1986]	5.1576	[5.1380, 5.1772]
110	1.7941	[1.7807, 1.8074]	11.7754	[11.7473, 11.8035]
120	0.4813	[0.4745, 0.4880]	20.4626	[20.4295, 20.4957]

Lookback Options				
Strike	Call	95% CI	Put	95% CI
80	39.1519	[39.1056, 39.1982]	3.8327	[3.8134, 3.8519]
90	29.1519	[29.1056, 29.1982]	8.9823	[8.9539, 9.0108]
100	19.1519	[19.1056, 19.1982]	17.2137	[17.1801, 17.2472]
110	11.1778	[11.1364, 11.2192]	27.2137	[27.1801, 27.2472]
120	6.1577	[6.1248, 6.1907]	37.2137	[37.1801, 37.2472]

Barrier Options (Up-barrier = 110, Down-barrier = 90)				
Strike	Up-In Call ($B = 110$)	95% CI	Down-Out Put ($B = 90$)	95% CI
80	19.8994	[19.8370, 19.9618]	0.0000	[0.0000, 0.0000]
90	14.2484	[14.1959, 14.3010]	0.0000	[0.0000, 0.0000]
100	9.3861	[9.3432, 9.4290]	0.1324	[0.1301, 0.1348]
110	5.6407	[5.6070, 5.6744]	0.8247	[0.8167, 0.8328]
120	3.1584	[3.1331, 3.1837]	2.2660	[2.2499, 2.2821]

Table 4: Option prices under the rough Heston model ($\alpha = 0.62$) using the INAR(∞) FFT-based simulator. All simulations use $\tau = 320$ and 500,000 paths.

in C++. For comparable accuracy, it runs faster than Euler-Maruyama baselines reported of [Richard et al. \(2023\)](#). The confidence intervals across option types remain tight.

4.2 Non-rough case: Heston model ($\alpha = 1$)

To validate our implementation, we conducted a comparative analysis between the rough Heston model implemented through INAR(∞) processes and the classical Heston model implemented via Euler-Maruyama discretization. Both models were calibrated with identical parameters as shown in Table 2, with the sole exception that $\alpha = 1$ was used for the classical Heston model to eliminate the roughness effect.

Since the closed-form solution for European option is available for the classical Heston model, we can directly compare the results from the INAR(∞) implementation to this benchmark. The results are given in Table 5. We choose $\tau = 320$, $S_0 = 100$ and $K = 80, 90, 100, 110, 120$ for all cases.

Table 5: Comparison of European option prices between closed-form solution and INAR(∞) implementations

Strike	Closed-form solution		INAR(∞)			
	Call	Put	Call	95% CI	Put	95% CI
80	21.8822	1.8822	21.8834	[21.8265, 21.9402]	1.8752	[1.8612, 1.8893]
90	14.6187	4.6187	14.6202	[14.5704, 14.6700]	4.6121	[4.5889, 4.6353]
100	9.0983	9.0983	9.0942	[9.0531, 9.1354]	9.0861	[9.0530, 9.1192]
110	5.2883	15.2883	5.2811	[5.2490, 5.3131]	15.2729	[15.2306, 15.3152]
120	2.8849	22.8849	2.8807	[2.8569, 2.9045]	22.8725	[22.8227, 22.9223]

The results for all other path-dependent option types under both models are presented in Tables 6–8. Since no closed-form solutions are available for these options, we use a carefully implemented Euler-Maruyama (E-M) scheme for the classical Heston model as a numerical benchmark. We provide the details of this scheme below for clarity and reproducibility.

The classical Heston model is simulated using a discrete time grid with a time step of $\Delta t = T/\tau = 1/320$. For each path, the dynamics of the asset price S and its variance V are discretized as follows:

$$S_{i+1} = S_i + rS_i\Delta t + \sqrt{\max(V_i, 0)}S_i\sqrt{\Delta t}Z_1,$$

$$V_{i+1} = \max\left(0, V_i + \gamma(\theta - V_i)\Delta t + \nu\sqrt{\max(V_i, 0)}\sqrt{\Delta t}Z_2\right),$$

where $i = 0, \dots, \tau - 1$, and (Z_1, Z_2) are correlated standard normal random variables with correlation ρ . This corresponds to a *diffusion-truncation scheme with terminal projection* (we use $\sqrt{\max(V_i, 0)}$ in the diffusion term and finally set $V_{i+1} \leftarrow \max\{V_{i+1}, 0\}$). To reduce the variance of the Monte Carlo estimates, we employ the antithetic variates technique. For each pair of generated random numbers (Z_1, Z_2) , we simulate a second path using $(-Z_1, -Z_2)$, and the final payoff for that simulation is the average of the payoffs from the two paths.

All benchmark prices from the Euler-Maruyama (E-M) scheme are computed with 500,000 paths (the same parameter settings for our INAR(∞) implementation); for the benchmarks we set $r = 0$.

Table 6: Comparison of Asian option prices between E-M method and INAR(∞) implementations

Strike	E-M method		INAR(∞)			
	Call	Put	Call	95% CI	Put	95% CI
80	20.1808	0.1808	20.1648	[20.1312, 20.1983]	0.1570	[0.1541, 0.1598]
90	11.3132	1.3132	11.2907	[11.2611, 11.3202]	1.2829	[1.2737, 1.2920]
100	4.9525	4.9525	4.9280	[4.9066, 4.9494]	4.9202	[4.9016, 4.9388]
110	1.6482	11.6482	1.6293	[1.6167, 1.6418]	11.6215	[11.5944, 11.6485]
120	0.4148	20.4148	0.4129	[0.4068, 0.4190]	20.4051	[20.3733, 20.4369]

Table 7: Comparison of Lookback option prices between E-M method and INAR(∞) implementations

Strike	E-M method		INAR(∞)			
	Call	Put	Call	95% CI	Put	95% CI
80	38.9575	3.4575	38.4427	[38.3980, 38.4874]	3.3807	[3.3630, 3.3984]
90	28.9575	8.4575	28.4427	[28.3980, 28.4874]	8.3549	[8.3279, 8.3819]
100	18.9575	16.9575	18.4427	[18.3980, 18.4874]	16.5222	[16.4900, 16.5544]
110	10.9575	26.9575	10.5300	[10.4903, 10.5697]	26.5222	[26.4900, 26.5544]
120	5.9575	36.9575	5.6523	[5.6211, 5.6834]	36.5222	[36.4900, 36.5544]

Table 8: Comparison of Barrier option prices between E-M method and INAR(∞) implementations (Up-barrier = 110, Down-barrier = 90)

Strike	E-M method	INAR(∞)	
	Up-In Call ($B = 110$)	Up-In Call ($B = 110$)	95% CI
80	19.5824	19.4397	[19.3788, 19.5006]
90	13.8715	13.8234	[13.7725, 13.8743]
100	8.9873	8.9829	[8.9416, 9.0241]
110	5.2883	5.2811	[5.2490, 5.3131]
120	2.8849	2.8807	[2.8569, 2.9045]

Strike	E-M method	INAR(∞)	
	Down-Out Put ($B = 90$)	Down-Out Put ($B = 90$)	95% CI
80	0.0000	0.0000	[0.0000, 0.0000]
90	0.0000	0.0000	[0.0000, 0.0000]
100	0.1308	0.1510	[0.1485, 0.1535]
110	0.8583	0.9308	[0.9222, 0.9393]
120	2.3692	2.5223	[2.5054, 2.5392]

The results demonstrate that our INAR(∞) implementation can effectively replicate the classical Heston model when $\alpha = 1$. European and Asian option prices show remarkable agreement between the two implementations, with differences typically less than 1%. For barrier options, the prices are also closely aligned, confirming that our approach correctly captures the path-dependent features present in exotic options.

For the classical Heston case, while our INAR(∞) approach requires 75.8623 seconds to compute all four types of options as compared to 3.6830 seconds for the Euler-Maruyama method, it offers a significant advantage in terms of model flexibility. Specifically, our framework can seamlessly accommodate both rough and classical volatility regimes without requiring changes to the underlying computational architecture. This unified treatment enables practitioners to efficiently explore different volatility assumptions within a single implementation framework before judging whether a rough or non-rough volatility model should be used, making it particularly valuable for comparative studies and model risk assessment.

4.3 Analysis of the Implied Volatility Surface

To validate the INAR(∞)-based simulator and to confront the empirical evidence behind the rough volatility, we examine the implied-volatility (IV) surface produced by our method. A key stylized fact is the short-maturity behavior of the at-the-money (ATM) skew: its magnitude increases as maturity shrinks and follows a power-law decay in T , a feature that classical models ($\alpha = 1$) typically understate.

Following the simulation setup in Section 4, we price European calls on a grid of maturities and log-moneyness $k = \log(K/S_0) \in [-0.2, 0.2]$. For each (T, k) we invert the Black-Scholes formula to obtain IVs. Figure 1 reports the resulting surface for the roughness specification ($\alpha = 0.62$), computed with the same path count and time step used throughout the study.

To make this stylized behavior more explicit, Figure 2 (a) traces the finite-difference ATM skew $k \mapsto \partial\sigma_{IV}/\partial k$ across maturities, while Figure 2 (b) shows the corresponding ATM implied volatilities. The skew steepens as maturity shrinks, whereas the ATM level follows a smoother upward term structure. Fitting a power law to the absolute skew (Figure 2, (c)) yields $|\text{skew}(T)| \approx cT^{-0.369}$ with $R^2 \approx 0.850$, closely matching the rough-volatility prediction $T^{H-1/2}$ implied by $\alpha = 0.62$ (so that $H = \alpha - \frac{1}{2} \approx 0.12$). This confirms that the simulator reproduces the near-expiry power-law decay that motivates rough volatility in practice.

Implied Volatility Surface (Comparison)

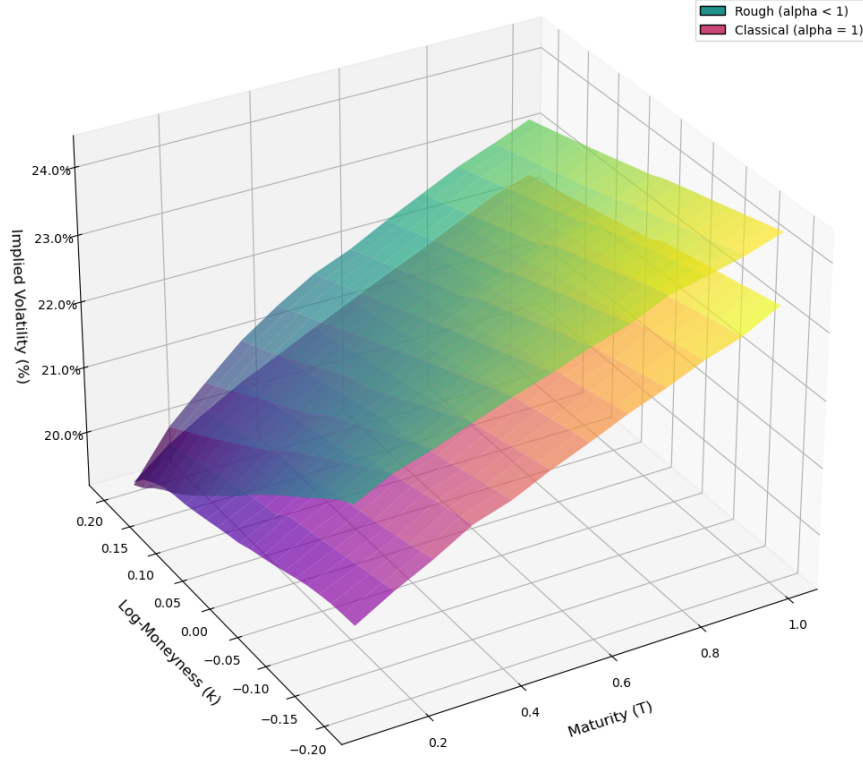


Figure 1: Implied-volatility surface produced by the roughness specification ($\alpha = 0.62$) with the INAR(∞)-FFT simulator. Grid: $T \in \{1/12, 2/12, 3/12, \dots, 1\}$, $k \in [-0.2, 0.2]$. All runs use a fixed time step ($\tau = 320$ per year) and 10^6 paths.

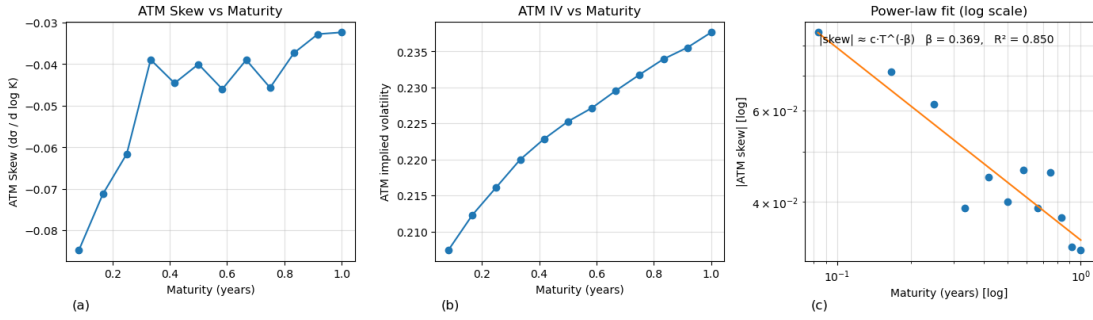


Figure 2: ATM diagnostics under the rough specification ($\alpha = 0.62$). Panel (a) shows the steepening skew toward short maturities; panel (b) reports the smoother ATM level term structure; panel (c) confirms the power-law decay $|\text{skew}(T)| \propto T^{H-1/2}$.

Overall, the INAR(∞) approximation reproduces the key empirical features of rough volatility within a unified simulation framework, providing numerical evidence that our micro-structural construction not only converges to rough Heston in theory, but also carries practical implications for options pricing.

5 Conclusion

We established a discrete microstructural foundation for the rough Heston model by proving that nearly-unstable, heavy-tailed *bivariate* cumulative INAR(∞) processes converge to the rough Heston dynamics. The construction delivers an explicit micro-to-macro link for both the variance and the log-price, including a transparent parameter mapping and the induced leverage structure.

On the computational side, we designed a divide-and-conquer FFT simulator that operates at the INAR prelimit rather than by directly time-stepping the Volterra SDE. The resulting Monte Carlo engine efficiently prices European and path-dependent contracts (arithmetic Asian, lookback, and barrier), produces tight confidence intervals across strikes and maturities, and continuously recovers the classical Heston benchmark as $\alpha \rightarrow 1$. The implied-volatility experiments confirm hallmark rough-volatility signatures, notably the steep short-maturity ATM skew with power-law decay.

Section 3 complements these contributions with *finite-horizon weak-error bounds*. For European and arithmetic Asian options, the INAR approximation attains

$$|C^{\text{INAR},\tau}(T, K) - C(T, K)| \leq C_{T,K} \tau^{\frac{1}{2}-\alpha}, \quad |AA^{\text{INAR},\tau}(T, K) - AA(T, K)| \leq C_{T,K} \tau^{\frac{1}{2}-\alpha},$$

while the microscopic layer contributes at most $C_{T,K} \tau^{-1/4}$ (Theorems 3.14 and 3.18). For fixed-strike lookback options we obtain

$$|LB^{\text{INAR},\tau}(T, K) - LB(T, K)| \leq C_{T,K} \left(\tau^{\frac{1}{4}-\frac{\alpha}{2}} + \tau^{-1/8} \sqrt{\log \tau} \right)$$

by Proposition 3.20 and Theorem 3.21.

Several directions appear promising. First, calibration and empirical testing on option panels (cross-sectional datasets across strikes and maturities, possibly over multiple dates) using the FFT-accelerated simulator; second, engineering improvements towards near- $\mathcal{O}(\tau \log \tau)$ complexity via overlap-save and kernel truncations; third, extensions to multi-asset settings and additional exotics; finally, applying the INAR-based microstructure to limit-order-book analytics and to risk management under rough volatility.

Acknowledgments

The authors would like to thank the AE and two anonymous referees for helpful comments. Lingjiong Zhu is partially supported by the grants NSF DMS-2053454, NSF DMS-2208303.

References

- Abi Jaber, E. and O. El Euch (2019). Multifactor approximation of rough volatility models. *SIAM J. Financ. Math.* 10(2), 309–349.
- Abi Jaber, E., M. Larsson, and S. Pulido (2019). Affine Volterra processes. *Ann. Appl. Probab.* 29(5), 3155–3200.
- Bayer, C. and S. Breneis (2024). Efficient option pricing in the rough Heston model using weak simulation schemes. *Quant. Finance* 24(9), 1247–1261.
- Bayer, C., P. Friz, and J. Gatheral (2016). Pricing under rough volatility. *Quant. Finance* 16(6), 887–904.
- Boyarchenko, S., M. de Innocentis, and S. Levendorskiĭ (2025). Fast reliable pricing and calibration of the rough Heston model. *arXiv preprint arXiv:2508.15080*.
- Callegaro, G., M. Grasselli, and G. Pages (2021). Fast hybrid schemes for fractional Riccati equations (rough is not so tough). *Math. Oper. Res.* 46(1), 221–254.

- Carr, P. and D. Madan (1999). Option valuation using the fast Fourier transform. *J. Comput. Finance* 2(4), 61–73.
- El Euch, O., M. Fukasawa, and M. Rosenbaum (2018). The microstructural foundations of leverage effect and rough volatility. *Financ. and Stoch.* 22, 241–280.
- El Euch, O. and M. Rosenbaum (2019). The characteristic function of rough Heston models. *Math. Finance* 29(1), 3–38.
- Fokianos, K. (2024). Multivariate count time series modelling. *Econom. and Stat.* 31, 100–116.
- Gatheral, J., T. Jaisson, and M. Rosenbaum (2018). Volatility is rough. *Quant. Finance* 18(6), 933–949.
- Gripenberg, G., S.-O. Londen, and O. Staffans (1990). *Volterra Integral and Functional Equations*, Volume 34. Cambridge University Press.
- Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. Academic press.
- Hawkes, A. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1), 83–90.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* 6(2), 327–343.
- Horst, U., W. Xu, and R. Zhang (2023). Convergence of heavy-tailed Hawkes processes and the microstructure of rough volatility. *arXiv preprint arXiv:2312.08784*.
- Huang, L. and M. Khabou (2023). Nonlinear Poisson autoregression and nonlinear Hawkes processes. *Stoch. Process. Appl.* 161, 201–241.
- Jacod, J. and A. Shiryaev (2013). *Limit Theorems for Stochastic Processes*, Volume 288. Springer Science & Business Media.
- Jaisson, T. and M. Rosenbaum (2015). Limit Theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.* 25(2), 600–631.
- Jaisson, T. and M. Rosenbaum (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Ann. Appl. Probab.* 25(2), 2860–2882.
- Keller-Ressel, M. and A. Majid (2020). A comparison principle between rough and non-rough Heston models—with applications to the volatility surface. *Quant. Finance* 20(6), 919–933.
- Kirchner, M. (2016). Hawkes and INAR (∞) processes. *Stoch. Process. Appl.* 126(8), 2494–2525.
- Liang, H., Z. Yang, and J. Gao (2017). Strong superconvergence of the Euler-Maruyama method for linear stochastic Volterra integral equations. *J. Comput. Appl. Math.* 317, 447–457.
- Ma, J., Z. Lu, and D. Chen (2023). Optimal reinsurance-investment with loss aversion under rough Heston model. *Quant. Finance* 23(1), 95–109.
- Ma, J. and H. Wu (2022). A fast algorithm for simulation of rough volatility models. *Quant. Finance* 22(3), 447–462.
- Olver, F. W. J., A. B. O. Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, and B. V. S. (eds.) (2023). Nist digital library of mathematical functions. Release 1.1.10, Chapter 25, §25.12 Polylogarithms.
- Richard, A., X. Tan, and F. Yang (2021). Discrete-time simulation of stochastic Volterra equations. *Stoch. Process. Appl.* 141, 109–138.
- Richard, A., X. Tan, and F. Yang (2023). On the discrete-time simulation of the rough Heston model. *SIAM J. Financ. Math.* 14(1), 223–249.

- Veraar, M. (2012). The stochastic Fubini theorem revisited. *Stochastics* 84(4), 543–551.
- Wang, Y., C. Cai, P. He, and Q. Wang (2024). Scaling limit of heavy tailed nearly unstable INAR (∞) processes and rough fractional diffusions. *arXiv preprint arXiv:2403.11773*.
- Xu, Y., L. Zhu, and H. Wang (2022). A self-and mutual-exciting model for discrete-time data: Case study on online money market fund. *Available at SSRN 3665436*.
- Yang, W., J. Ma, and Z. Cui (2025). A general valuation framework for rough stochastic local volatility models and applications. *Eur. J. Oper. Res.* 322(1), 307–324.

A Proofs of Section 2

Throughout the proofs in the appendix, we use the following asymptotic notation. We write $f(\tau) \sim g(\tau)$ to denote asymptotic equivalence as $\tau \rightarrow \infty$, i.e., $\lim_{\tau \rightarrow \infty} f(\tau)/g(\tau) = 1$. We use $f(\tau) \lesssim g(\tau)$ to indicate that $f(\tau) \leq Cg(\tau)$ for some positive constant C that is independent of τ for all sufficiently large τ . We write $x_\tau \asymp y_\tau$ to mean that there exist constants $0 < c < C < \infty$ such that $cy_\tau \leq x_\tau \leq Cy_\tau$ for all sufficiently large τ .

A.1 Proof of Lemma 2.5

Fix $\alpha \in (0, 1)$, $\gamma > 0$ and $z > 0$. Let $(\varphi_n)_{n \geq 1}$ be a base kernel with $\sum_{n \geq 1} \varphi_n = 1$ and tail

$$\bar{\varphi}(n) := \sum_{k > n} \varphi_k \sim \frac{1}{\Gamma(1 - \alpha)} n^{-\alpha} \quad (n \rightarrow \infty).$$

For each $\tau > 0$ pick $a_\tau \in (0, 1)$ so that $\tau^\alpha(1 - a_\tau) \rightarrow \gamma$, and set $\varphi_n^\tau := a_\tau \varphi_n$. Let $(G_i)_{i \geq 1}$ be i.i.d. \mathbb{N} -valued random variables with $\mathbb{P}(G_1 = n) = \varphi_n$. For $n \geq 1$ define the (discrete) renewal kernel

$$\psi_n^\tau := \sum_{k \geq 1} (\varphi^\tau)_n^{*k} = \sum_{k \geq 1} a_\tau^k (\varphi^{*k})_n.$$

Define the rescaled renewal density

$$\zeta^\tau(t) := \frac{(1 - a_\tau)}{a_\tau} \tau \psi_{\lfloor t\tau \rfloor}^\tau, \quad t \geq 0.$$

A one-line check shows normalization:

$$\int_0^\infty \zeta^\tau(t) dt = \frac{1 - a_\tau}{a_\tau} \sum_{n=1}^\infty \psi_n^\tau = \frac{1 - a_\tau}{a_\tau} \cdot \frac{a_\tau}{1 - a_\tau} = 1.$$

Consider the Laplace transform $\hat{\zeta}^\tau(z) := \int_0^\infty e^{-zt} \zeta^\tau(t) dt$. Since ζ^τ is piecewise constant on $[n/\tau, (n+1)/\tau)$, we obtain

$$\begin{aligned} \hat{\zeta}^\tau(z) &= \frac{(1 - a_\tau)}{a_\tau} \tau \sum_{n \geq 1} \psi_n^\tau \int_{n/\tau}^{(n+1)/\tau} e^{-zt} dt \\ &= \frac{(1 - a_\tau)}{a_\tau} \tau \sum_{n \geq 1} \psi_n^\tau \frac{e^{-zn/\tau} - e^{-z(n+1)/\tau}}{z} \\ &= \frac{(1 - a_\tau)}{a_\tau} \frac{1 - e^{-z/\tau}}{z} \tau \sum_{n \geq 1} \psi_n^\tau e^{-zn/\tau}. \end{aligned}$$

Write the generating function $\hat{\varphi}(s) := \sum_{n \geq 1} \varphi_n e^{-sn}$ (so that $\hat{\varphi}(s) = G_\varphi(e^{-s})$). Then

$$\sum_{n \geq 1} \psi_n^\tau e^{-zn/\tau} = \frac{G_{\varphi^\tau}(e^{-z/\tau})}{1 - G_{\varphi^\tau}(e^{-z/\tau})} = \frac{a_\tau \hat{\varphi}(z/\tau)}{1 - a_\tau \hat{\varphi}(z/\tau)}.$$

Therefore

$$\hat{\zeta}^\tau(z) = (1 - a_\tau) \frac{1 - e^{-z/\tau}}{z} \tau \frac{\hat{\varphi}(z/\tau)}{1 - a_\tau \hat{\varphi}(z/\tau)}.$$

Since $\tau(1 - e^{-z/\tau})/z \rightarrow 1$ as $\tau \rightarrow \infty$, it follows that

$$\lim_{\tau \rightarrow \infty} \hat{\zeta}^\tau(z) = \lim_{\tau \rightarrow \infty} \frac{(1 - a_\tau) \hat{\varphi}(z/\tau)}{1 - a_\tau \hat{\varphi}(z/\tau)}.$$

It remains to expand $\hat{\varphi}(s)$ as $s \downarrow 0$. By Abel summation (summation by parts),

$$\hat{\varphi}(s) = \sum_{n \geq 1} \varphi_n e^{-sn} = 1 - (1 - e^{-s}) \sum_{n \geq 0} e^{-sn} \bar{\varphi}(n).$$

Using $\bar{\varphi}(n) \sim \Gamma(1-\alpha)^{-1} n^{-\alpha}$ and the classical asymptotic $\sum_{n \geq 1} e^{-sn} n^{-\alpha} \sim \Gamma(1-\alpha) s^{\alpha-1}$ as $s \downarrow 0$, we obtain

$$(1 - e^{-s}) \sum_{n \geq 0} e^{-sn} \bar{\varphi}(n) \sim s^\alpha, \quad \hat{\varphi}(s) = 1 - s^\alpha + o(s^\alpha).$$

With $s = z/\tau$,

$$1 - a_\tau \hat{\varphi}(z/\tau) = (1 - a_\tau) + a_\tau (z/\tau)^\alpha + o(\tau^{-\alpha}).$$

Hence

$$\lim_{\tau \rightarrow \infty} \hat{\zeta}^\tau(z) = \lim_{\tau \rightarrow \infty} \frac{(1 - a_\tau)(1 - (z/\tau)^\alpha + o(\tau^{-\alpha}))}{(1 - a_\tau) + a_\tau (z/\tau)^\alpha + o(\tau^{-\alpha})} = \frac{1}{1 + z^\alpha/\gamma} = \frac{\gamma}{\gamma + z^\alpha}.$$

The limit function is continuous on $[0, \infty)$ and equals 1 at $z = 0$. By the continuity theorem for Laplace transforms of probability measures on $[0, \infty)$,

$$\zeta^\tau(t) dt \Rightarrow f_{\alpha, \gamma}(t) dt, \quad f_{\alpha, \gamma}(t) = \gamma t^{\alpha-1} E_{\alpha, \alpha}(-\gamma t^\alpha), \quad t > 0,$$

where $E_{\alpha, \beta}$ is the two-parameter Mittag-Leffler function. This completes the proof.

A.2 Proof of Proposition 2.6

For each integer $\tau \in \{1, 2, 3, \dots\}$, we work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$. The law of the process depends on τ through its parameters, but the underlying probability space is held fixed.

C-tightness of \mathbf{Y}^τ and Λ^τ :

Since $Y_0^\tau = 0$, the condition $\sup_{\tau \geq 1} \mathbb{E}[|Y_0^\tau|^\beta] = 0 < \infty$ is trivially satisfied. We apply Lemma 2.7 to the sequence of càdlàg processes $(Y_t^\tau)_{t \in [0, 1]}$. We need to verify the two conditions of the lemma.

(i) Given a constant $\theta > 2$,

$$\begin{aligned} \sup_{k=0, 1, \dots, \lfloor \tau^\theta \rfloor - 1} \sup_{h \in [0, 1/\tau^\theta]} |\Delta_h Y_{k/\tau^\theta}^{\tau, +}| &= \sup_{k=0, 1, \dots, \lfloor \tau^\theta \rfloor - 1} \left| Y_{(k+1)/\tau^\theta}^{\tau, +} - Y_{k/\tau^\theta}^{\tau, +} \right| \\ &= \frac{1 - a_\tau}{\tau^\alpha \mu} \sup_{k=0, 1, \dots, \lfloor \tau^\theta \rfloor - 1} \sum_{s=\lfloor k/\tau^{\theta-1} \rfloor + 1}^{\lfloor (k+1)/\tau^{\theta-1} \rfloor} X_s^{\tau, +} \\ &= \frac{1 - a_\tau}{\tau^\alpha \mu} \sup_{k=0, 1, \dots, \lfloor \tau^\theta \rfloor - 1} X_{\lfloor \frac{k+1}{\tau^{\theta-1}} \rfloor}^{\tau, +}. \end{aligned}$$

We need to prove for any $\epsilon > 0$,

$$\mathbb{P} \left(\frac{1 - a_\tau}{\tau^\alpha \mu} \sup_{k=0, 1, \dots, \lfloor \tau^\theta \rfloor - 1} X_{\lfloor \frac{k+1}{\tau^{\theta-1}} \rfloor}^{\tau, +} > \epsilon \right) \rightarrow 0 \quad (\text{A.1})$$

as $\tau \rightarrow \infty$. In fact, since there are at most $\lfloor \tau \rfloor$ integers in $\{\lfloor \frac{k+1}{\tau^{\theta-1}} \rfloor : k = 0, 1, \dots, \lfloor \tau^\theta \rfloor - 1\}$, from the Markov inequality,

$$\begin{aligned} \mathbb{P} \left(\frac{1 - a_\tau}{\tau^\alpha \mu} \sup_{k=0, 1, \dots, \lfloor \tau^\theta \rfloor - 1} X_{\lfloor \frac{k+1}{\tau^{\theta-1}} \rfloor}^{\tau, +} > \epsilon \right) &= \mathbb{P} \left(\sup_{k=0, 1, \dots, \lfloor \tau^\theta \rfloor - 1} X_{\lfloor \frac{k+1}{\tau^{\theta-1}} \rfloor}^{\tau, +} > \frac{\tau^\alpha \mu}{1 - a_\tau} \epsilon \right) \\ &\leq \sum_{s=0}^{\lfloor \tau \rfloor} \frac{\mathbb{E}[(X_s^{\tau, +})^2]}{\left(\frac{\tau^\alpha \mu}{1 - a_\tau} \epsilon \right)^2} \end{aligned}$$

$$= \sum_{s=0}^{\lfloor \tau \rfloor} \frac{\mathbb{E}[\lambda_s^{\tau,+}] + \mathbb{E}[(\lambda_s^{\tau,+})^2]}{\left(\frac{\tau^\alpha \mu}{1-a_\tau} \epsilon\right)^2}.$$

Fix $t \in (0, 1]$ and set $s = \lfloor \tau t \rfloor$ (so that $s/\tau \rightarrow t$). We now estimate the magnitude of $\mathbb{E}[\lambda_s^{\tau,+}]$ and $\mathbb{E}[(\lambda_s^{\tau,+})^2]$, uniformly in t . We calculate them step by step.

- (a) We begin by deriving the asymptotic order of $\hat{\mu}_\tau(s)$ for s on the order of τ . Let $s = \lfloor \tau t \rfloor$ for a fixed $t \in (0, 1)$. Recalling the definition of $\hat{\mu}_\tau(n)$ from Definition 2.3:

$$\hat{\mu}_\tau(\lfloor \tau t \rfloor) = \mu_\tau + \xi \mu_\tau \left(\frac{1}{1-a_\tau} \left(1 - \sum_{s=1}^{\lfloor \tau t \rfloor - 1} \varphi_s^\tau \right) - \sum_{s=1}^{\lfloor \tau t \rfloor - 1} \varphi_s^\tau \right).$$

We analyze the terms inside the parenthesis. First, note that $\varphi_s^\tau = a_\tau \varphi_s$ and $\sum_{s=1}^\infty \varphi_s = 1$. Thus,

$$1 - \sum_{s=1}^{\lfloor \tau t \rfloor - 1} \varphi_s^\tau = 1 - a_\tau \sum_{s=1}^{\lfloor \tau t \rfloor - 1} \varphi_s = 1 - a_\tau \left(1 - \sum_{s=\lfloor \tau t \rfloor}^\infty \varphi_s \right) = (1 - a_\tau) + a_\tau \sum_{s=\lfloor \tau t \rfloor}^\infty \varphi_s.$$

From the heavy-tailed property of the kernel (see Assumption 2.1), we have $\sum_{s=n}^\infty \varphi_s \sim \frac{1}{\Gamma(1-\alpha)n^\alpha}$ for large n . Setting $n = \lfloor \tau t \rfloor$, we get $\sum_{s=\lfloor \tau t \rfloor}^\infty \varphi_s \sim \frac{1}{\Gamma(1-\alpha)(\tau t)^\alpha}$. Substituting this back, the first term in the parenthesis becomes:

$$\frac{1}{1-a_\tau} \left((1-a_\tau) + a_\tau \sum_{s=\lfloor \tau t \rfloor}^\infty \varphi_s \right) = 1 + \frac{a_\tau}{1-a_\tau} \sum_{s=\lfloor \tau t \rfloor}^\infty \varphi_s.$$

Since $a_\tau = 1 - \gamma\tau^{-\alpha}$, we have $1 - a_\tau = \gamma\tau^{-\alpha}$. The term above is asymptotically equivalent to:

$$1 + \frac{1}{\gamma\tau^{-\alpha}} \frac{1}{\Gamma(1-\alpha)(\tau t)^\alpha} = 1 + \frac{1}{\gamma\Gamma(1-\alpha)t^\alpha}.$$

The second term in the parenthesis is $\sum_{s=1}^{\lfloor \tau t \rfloor - 1} \varphi_s^\tau = a_\tau \sum_{s=1}^{\lfloor \tau t \rfloor - 1} \varphi_s \rightarrow 1$ as $\tau \rightarrow \infty$.

Combining these results, we find the asymptotic behavior of $\hat{\mu}_\tau(\lfloor \tau t \rfloor)$:

$$\hat{\mu}_\tau(\lfloor \tau t \rfloor) \sim \mu_\tau + \xi \mu_\tau \left(\left(1 + \frac{1}{\gamma\Gamma(1-\alpha)t^\alpha} \right) - 1 \right) = \mu_\tau \left(1 + \frac{\xi}{\gamma\Gamma(1-\alpha)t^\alpha} \right).$$

Since $\mu_\tau = \mu\tau^{\alpha-1}$ and the term in the parenthesis is a constant with respect to τ , we conclude that:

$$\hat{\mu}_\tau(\lfloor \tau t \rfloor) \sim \mu\tau^{\alpha-1} \left(1 + \frac{\xi}{\gamma\Gamma(1-\alpha)t^\alpha} \right) = \mathcal{O}(\tau^{\alpha-1}).$$

This establishes that for large τ , $\hat{\mu}_\tau(s)$ is of the order $\tau^{\alpha-1}$ when $s \asymp \tau$.

- (b) From (a), for $t \in (0, 1]$ set $n = \lfloor \tau t \rfloor$. Using the bound from (a),

$$\hat{\mu}_\tau(s) \lesssim \mu\tau^{\alpha-1} \left(1 - \xi + \frac{\xi}{\gamma\Gamma(1-\alpha)} \left(\frac{s}{\tau} \right)^{-\alpha} \right), \quad 1 \leq s \leq n-1,$$

we estimate the discrete convolution:

$$\sum_{s=1}^{n-1} \psi_{n-s}^\tau \hat{\mu}_\tau(s) \lesssim \mu\tau^{\alpha-1} \left[(1-\xi) \sum_{j=1}^{n-1} \psi_j^\tau + \frac{\xi}{\gamma\Gamma(1-\alpha)} \sum_{j=1}^{n-1} \psi_j^\tau \left(\frac{\tau}{n-j} \right)^\alpha \right],$$

where we put $j = n - s$. Define $\zeta^\tau(u) := \frac{1-a_\tau}{a_\tau} \tau \psi_{\lfloor \tau u \rfloor}^\tau$. By Lemma 2.5, $\zeta^\tau(u) du \Rightarrow f_{\alpha, \gamma}(u) du$ on $[0, \infty)$. Hence,

$$\frac{1-a_\tau}{a_\tau} \cdot \frac{1}{\tau} \sum_{j=1}^{n-1} \psi_j^\tau = \frac{1}{\tau} \sum_{j=1}^{n-1} \zeta^\tau\left(\frac{j}{\tau}\right) \longrightarrow \int_0^t f_{\alpha, \gamma}(u) du \leq 1,$$

and, writing $n - j = \tau(t - \frac{j}{\tau})$,

$$\frac{1-a_\tau}{a_\tau} \cdot \frac{1}{\tau} \sum_{j=1}^{n-1} \psi_j^\tau \left(\frac{\tau}{n-j}\right)^\alpha = \frac{1}{\tau} \sum_{j=1}^{n-1} \zeta^\tau\left(\frac{j}{\tau}\right) \left(t - \frac{j}{\tau}\right)^{-\alpha} \longrightarrow \int_0^t f_{\alpha, \gamma}(u) (t-u)^{-\alpha} du,$$

which is finite for $\alpha \in (0, 1)$. Therefore,

$$\sum_{s=1}^{\lfloor \tau t \rfloor - 1} \psi_{\lfloor \tau t \rfloor - s}^\tau \hat{\mu}_\tau(s) \lesssim \mu \tau^{\alpha-1} \cdot \frac{a_\tau}{1-a_\tau} \cdot C_t \lesssim \mu \tau^{2\alpha-1},$$

with C_t bounded uniformly in $t \in (0, 1]$.

(c) From (2.2), it follows that for $t \in (0, 1)$,

$$\mathbb{E} \left[\lambda_{\lfloor \tau t \rfloor}^{\tau, +} \right] = \hat{\mu}_\tau(\lfloor \tau t \rfloor) + \sum_{s=1}^{\lfloor \tau t \rfloor - 1} \psi_{\lfloor \tau t \rfloor - s}^\tau \hat{\mu}_\tau(s) \lesssim \tau^{2\alpha-1}. \quad (\text{A.2})$$

(d) Use (2.2), define

$$\kappa_n^{\tau, +} = \frac{1}{1+\beta} \sum_{s=1}^{n-1} \psi_{n-s}^\tau (M_s^{\tau, +} - M_{s-1}^{\tau, +} + \beta M_s^{\tau, -} - \beta M_{s-1}^{\tau, -}). \quad (\text{A.3})$$

Denote

$$f_{\alpha, \gamma}(t) = \gamma t^{\alpha-1} E_{\alpha, \alpha}(-\gamma t^\alpha),$$

where

$$E_{\alpha, \beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}$$

is the Mittag-Leffler function. We use the weak convergence of the renewal kernel, which we establish in Lemma 2.5 in the appendix. This result states that the measure associated with the function $\frac{(1-a_\tau)}{a_\tau} \tau \psi_{\lfloor \tau \cdot \rfloor}^\tau$ converges weakly to the measure with density $f_{\alpha, \gamma}$. This allows us to approximate the discrete convolution with a continuous one. Hence, from (A.2), for $t \in [0, 1]$, we compute

$$\begin{aligned} \mathbb{E} \left[\left(\kappa_{\lfloor t\tau \rfloor}^{\tau, +} \right)^2 \right] &= \frac{1}{(1+\beta)^2} \sum_{s=1}^{\lfloor t\tau \rfloor - 1} \left(\psi_{\lfloor t\tau \rfloor - s}^\tau \right)^2 \mathbb{E} \left[\left((X_s^{\tau, +} - \lambda_s^{\tau, +}) + \beta (X_s^{\tau, -} - \lambda_s^{\tau, -}) \right)^2 \right] \\ &\leq \frac{1}{(1+\beta)^2} \sum_{s=1}^{\lfloor t\tau \rfloor - 1} \left(\psi_{\lfloor t\tau \rfloor - s}^\tau \right)^2 \left(2\mathbb{E} \left[(X_s^{\tau, +} - \lambda_s^{\tau, +})^2 \right] + 2\beta^2 \mathbb{E} \left[(X_s^{\tau, -} - \lambda_s^{\tau, -})^2 \right] \right) \\ &= \frac{2}{(1+\beta)^2} \sum_{s=1}^{\lfloor t\tau \rfloor - 1} \left(\psi_{\lfloor t\tau \rfloor - s}^\tau \right)^2 \left(\mathbb{E} \left[\lambda_s^{\tau, +} \right] + \beta^2 \mathbb{E} \left[\lambda_s^{\tau, -} \right] \right) \\ &= \frac{2(1+\beta^2)}{(1+\beta)^2} \sum_{s=1}^{\lfloor t\tau \rfloor - 1} \left(\psi_{\lfloor t\tau \rfloor - s}^\tau \right)^2 \mathbb{E} \left[\lambda_s^{\tau, +} \right]. \end{aligned}$$

Using the result from a previous step that $\mathbb{E}[\lambda_s^{\tau,+}] = \mathcal{O}(\tau^{2\alpha-1})$ uniformly for $s \leq \lfloor t\tau \rfloor$, which implies the existence of a positive constant C (independent of τ and s) such that $\mathbb{E}[\lambda_s^{\tau,+}] \leq C\tau^{2\alpha-1}$ for all $s \leq \lfloor t\tau \rfloor$ and sufficiently large τ . As a result, we can bound the expression above:

$$\begin{aligned} \mathbb{E} \left[\left(\kappa_{\lfloor t\tau \rfloor}^{\tau,+} \right)^2 \right] &\lesssim \tau^{2\alpha-1} \sum_{s=1}^{\lfloor t\tau \rfloor-1} \left(\psi_{\lfloor t\tau \rfloor-s}^{\tau} \right)^2 = \tau^{2\alpha-1} \sum_{j=1}^{\lfloor t\tau \rfloor-1} \left(\psi_j^{\tau} \right)^2 \\ &\lesssim \tau^{2\alpha-1} \sum_{j=1}^{\lfloor t\tau \rfloor-1} \left(\frac{a_{\tau}}{(1-a_{\tau})^{\tau}} f_{\alpha,\gamma} \left(\frac{j}{\tau} \right) \right)^2 \\ &\lesssim \tau^{2\alpha-1} \left(\frac{\tau^{\alpha}}{\gamma} \right)^2 \frac{1}{\tau} \sum_{j=1}^{\lfloor t\tau \rfloor-1} \left(\left(\frac{j}{\tau} \right)^{\alpha-1} E_{\alpha,\alpha} \left(-\gamma \left(\frac{j}{\tau} \right)^{\alpha} \right) \right)^2 \\ &\lesssim \tau^{4\alpha-2}, \quad \text{uniformly in } t \in [0, 1]. \end{aligned}$$

Since $\alpha \in (1/2, 1)$, the exponent $2\alpha-2$ is greater than -1 , which ensures that the integral converges to a finite positive constant. Thus, we have shown that $\mathbb{E} \left[\left(\kappa_{\lfloor t\tau \rfloor}^{\tau,+} \right)^2 \right] = \mathcal{O}(\tau^{4\alpha-2})$. Hence, there exists a constant C (independent of t and τ for τ large) such that

$$\mathbb{E} \left[\left(\kappa_{\lfloor t\tau \rfloor}^{\tau,+} \right)^2 \right] \leq C\tau^{4\alpha-2}.$$

From (A.2) and the decomposition $\lambda_s^{\tau,+} = (\lambda_s^{\tau,+} - \kappa_s^{\tau,+}) + \kappa_s^{\tau,+}$,

$$\mathbb{E} [(\lambda_s^{\tau,+})^2] \leq 2(\lambda_s^{\tau,+} - \kappa_s^{\tau,+})^2 + 2\mathbb{E} [(\kappa_s^{\tau,+})^2] \lesssim \tau^{4\alpha-2},$$

uniformly for $s \leq \lfloor t\tau \rfloor$, $t \in [0, 1]$, and large τ .

As a result, using $\frac{\tau^{\alpha}\mu}{1-a_{\tau}} \gtrsim \tau^{2\alpha}$ and the uniform bounds above,

$$\sum_{s=0}^{\lfloor \tau \rfloor} \frac{\mathbb{E}[\lambda_s^{\tau}] + \mathbb{E}[(\lambda_s^{\tau})^2]}{\left(\frac{\tau^{\alpha}\mu}{1-a_{\tau}} \epsilon \right)^2} \lesssim \tau^{-1} \rightarrow 0.$$

as $\tau \rightarrow \infty$. This implies $\sup_{k=0,1,\dots,\lfloor \tau^{\theta} \rfloor-1} \sup_{h \in [0,1/\tau^{\theta}]} \left| \Delta_h Y_{k/\tau^{\theta}}^{\tau} \right| \rightarrow 0$ in probability as $\tau \rightarrow \infty$. This implies the first condition in Lemma 2.7 is satisfied.

(ii) Set $p = 2$, $a_1 = 0$, $b_1 = 2$, $\rho = 2$, for $t \in (0, 1]$, from (A.2) and a fundamental inequality

$$\left(\sum_{i=1}^n x_i \right)^2 \leq n \sum_{i=1}^n x_i^2,$$

we obtain

$$\begin{aligned} \mathbb{E} \left[\left| \Delta_h Y_t^{\tau,+} \right|^2 \right] &= \mathbb{E} \left[\left(\frac{1-a_{\tau}}{\tau^{\alpha}\mu} \right)^2 \left(\sum_{i=\lfloor \tau t \rfloor}^{\lfloor \tau(t+h) \rfloor} X_i^{\tau,+} \right)^2 \right] \\ &\leq \left(\frac{1-a_{\tau}}{\tau^{\alpha}\mu} \right)^2 (\lfloor \tau(t+h) \rfloor - \lfloor \tau t \rfloor + 1) \sum_{i=\lfloor \tau t \rfloor}^{\lfloor \tau(t+h) \rfloor} \mathbb{E} \left[(X_i^{\tau,+})^2 \right]. \end{aligned}$$

We can estimate the order of $\mathbb{E}[(X_s^{\tau,+})^2]$ for sufficiently large τ , uniformly in $t \in (0, 1]$ with $s = \lfloor \tau t \rfloor$. Conditionally on the past \mathcal{F}_{s-1} , we have $X_s^{\tau,+} | \mathcal{F}_{s-1} \sim \text{Poisson}(\lambda_s^{\tau,+})$, hence

$$\mathbb{E} [(X_s^{\tau,+})^2] = \mathbb{E} [\text{Var}(X_s^{\tau,+} | \mathcal{F}_{s-1}) + (\mathbb{E}[X_s^{\tau,+} | \mathcal{F}_{s-1}])^2] = \mathbb{E}[\lambda_s^{\tau,+}] + \mathbb{E}[(\lambda_s^{\tau,+})^2].$$

In the nearly-unstable regime $1 - a_\tau \sim \gamma\tau^{-\alpha}$ and with s of order τ , standard resolvent arguments yield $\mathbb{E}[\lambda_s^{\tau,+}] \asymp \mu_\tau/(1 - a_\tau)$ with $\mu_\tau = \mu\tau^{\alpha-1}$; thus $\lambda_s^{\tau,+} \lesssim \tau^{2\alpha-1}$ and, uniformly for s of order τ ,

$$\mathbb{E}[(X_s^{\tau,+})^2] \lesssim \tau^{4\alpha-2}.$$

Consequently,

$$\begin{aligned} & \left(\frac{1-a_\tau}{\mu\tau^\alpha}\right)^2 (\lfloor \tau(t+h) \rfloor - \lfloor \tau t \rfloor + 1) \sum_{i=\lfloor \tau t \rfloor}^{\lfloor \tau(t+h) \rfloor} \mathbb{E}[(X_i^{\tau,+})^2] \\ & \sim \left(\frac{1-a_\tau}{\mu\tau^\alpha}\right)^2 (\lfloor \tau(t+h) \rfloor - \lfloor \tau t \rfloor + 1)^2 \tau^{4\alpha-2} \sim h^2, \end{aligned}$$

as $\tau \rightarrow \infty$. This verifies the second condition in Lemma 2.7.

C-tightness of \mathbf{Z}^τ and Characterization of the Limit: The proof of tightness for the \mathbf{Z}^τ sequence and the characterization of its limit involve two main parts: establishing the convergence of its quadratic variation and then identifying the properties of the limit point.

1. *Convergence of the Quadratic Variation.* We aim to show that for the ‘+’ component, $[Z^{\tau,+}, Z^{\tau,+}]_t - Y_t^{\tau,+} \rightarrow 0$ in probability as $\tau \rightarrow \infty$. The argument for the ‘-’ component is identical. This is achieved by proving that both $[Z^{\tau,+}, Z^{\tau,+}]_t$ and $Y_t^{\tau,+}$ converge to the same limit as the predictable quadratic variation, $\Lambda_t^{\tau,+}$. This is done in two sub-steps:

(a) We first show that the difference between the quadratic variation and the predictable quadratic variation vanishes. Let $C_\tau^2 = \frac{\tau^\alpha \mu}{1-a_\tau}$ and $D_\tau = \frac{1-a_\tau}{\tau^\alpha \mu}$. The process $U_n^{\tau,+} := [Z^{\tau,+}, Z^{\tau,+}]_{n/\tau} - \langle Z^{\tau,+}, Z^{\tau,+} \rangle_{n/\tau}$ is a martingale. We prove that it converges to zero in L^2 , which implies convergence in probability.

$$\begin{aligned} \mathbb{E}[(U_{\lfloor t\tau \rfloor}^{\tau,+})^2] &= \mathbb{E}\left[\left(D_\tau^2 C_\tau^2 \sum_{s=1}^{\lfloor t\tau \rfloor} ((X_s^{\tau,+} - \lambda_s^{\tau,+})^2 - \lambda_s^{\tau,+})\right)^2\right] \\ &= D_\tau^2 \sum_{s=1}^{\lfloor t\tau \rfloor} \mathbb{E}\left[\left((X_s^{\tau,+} - \lambda_s^{\tau,+})^2 - \lambda_s^{\tau,+}\right)^2\right] \\ &= D_\tau^2 \sum_{s=1}^{\lfloor t\tau \rfloor} \mathbb{E}\left[\text{Var}\left((X_s^{\tau,+} - \lambda_s^{\tau,+})^2 | \mathcal{F}_{s-1}\right)\right] \\ &= D_\tau^2 \sum_{s=1}^{\lfloor t\tau \rfloor} \mathbb{E}\left[2(\lambda_s^{\tau,+})^2 + \lambda_s^{\tau,+}\right]. \end{aligned}$$

Using the uniform bounds $\mathbb{E}[\lambda_s^{\tau,+}] \lesssim \tau^{2\alpha-1}$ and $\mathbb{E}[(\lambda_s^{\tau,+})^2] \lesssim \tau^{4\alpha-2}$ derived earlier, the sum is bounded by $\mathcal{O}(\tau \cdot \tau^{4\alpha-2}) = \mathcal{O}(\tau^{4\alpha-1})$. The pre-factor $D_\tau^2 \sim (\tau^{-\alpha})^2 = \tau^{-4\alpha}$. Thus,

$$\mathbb{E}[(U_{\lfloor t\tau \rfloor}^{\tau,+})^2] \lesssim \tau^{-4\alpha} \cdot \tau^{4\alpha-1} = \tau^{-1} \rightarrow 0.$$

By Doob’s inequality, this implies $\sup_{t \in [0,1]} |[Z^{\tau,+}, Z^{\tau,+}]_t - \langle Z^{\tau,+}, Z^{\tau,+} \rangle_t| \rightarrow 0$ in probability.

(b) We show that $\langle Z^{\tau,+}, Z^{\tau,+} \rangle_t$ and $Y_t^{\tau,+}$ converge to the same limit. As shown previously, we have the identity $\langle Z^{\tau,+}, Z^{\tau,+} \rangle_t = \Lambda_t^{\tau,+}$. Furthermore, in Appendix A.3, we prove that $\sup_{t \in [0,1]} |Y_t^{\tau,+} - \Lambda_t^{\tau,+}| \rightarrow 0$ in probability. Therefore, $\langle Z^{\tau,+}, Z^{\tau,+} \rangle_t$ and $Y_t^{\tau,+}$ converge to the same limit process.

Combining (a) and (b), we conclude that $[Z^{\tau,+}, Z^{\tau,+}]_t - Y_t^{\tau,+} \rightarrow 0$ in probability.

2. *Characterization of the Limit.* Let (\mathbf{Y}, \mathbf{Z}) be any limit point of the tight sequence $(\mathbf{Y}^\tau, \mathbf{Z}^\tau)$. From the convergence of the quadratic variation, we have $[\mathbf{Z}, \mathbf{Z}] = \text{diag}(\mathbf{Y})$.

The tightness of $(Z^{\tau,+})$ and the convergence of its quadratic variation to a continuous process Y^+ imply that the limit Z^+ is a continuous martingale (e.g., Proposition VI-3.26 of [Jacod and Shiryaev \(2013\)](#)). The same argument applies to the ‘ $-$ ’ component.

Finally, since the martingale increments $(X_s^{\tau,+} - \lambda_s^{\tau,+})$ and $(X_s^{\tau,-} - \lambda_s^{\tau,-})$ are uncorrelated given \mathcal{F}_{s-1} (as they are derived from conditionally independent Poisson variables), their cross-variation $[Z^{\tau,+}, Z^{\tau,-}]_t$ converges to zero. This ensures that the limit process $\mathbf{Z} = (Z^+, Z^-)^\top$ has diagonal quadratic variation, i.e., $[Z^+, Z^-]_t = 0$.

This completes the proof of Proposition 2.6.

A.3 Proof of Lemma 2.8

Without loss of generality, it suffices to verify that

$$\sup_{t \in [0,1]} |\Lambda_t^{\tau,+} - Y_t^{\tau,+}| \rightarrow 0 \text{ in probability}$$

as $\tau \rightarrow \infty$. Indeed, using (A.2) and applying Doob’s inequality to $M^{\tau,+}$, we obtain

$$\begin{aligned} \sup_{t \in [0,1]} |Y_t^{\tau,+} - \Lambda_t^{\tau,+}| &\leq C \left(\frac{1-a_\tau}{\tau^\alpha} \right)^2 \mathbb{E} \left[\left(M_{\lfloor \tau \rfloor}^{\tau,+} \right)^2 \right] \\ &= C \left(\frac{1-a_\tau}{\tau^\alpha} \right)^2 \mathbb{E} \left[\sum_{s=1}^{\lfloor \tau \rfloor} (X_s^{\tau,+} - \lambda_s^{\tau,+})^2 + 2 \sum_{1 \leq i < j \leq \lfloor \tau \rfloor} (X_i^{\tau,+} - \lambda_i^{\tau,+}) (X_j^{\tau,+} - \lambda_j^{\tau,+}) \right] \\ &= C \left(\frac{1-a_\tau}{\tau^\alpha} \right)^2 \sum_{s=1}^{\lfloor \tau \rfloor} \mathbb{E}[\lambda_s^{\tau,+}] \\ &\sim C \tau^{-2\alpha}. \end{aligned}$$

This completes the proof.

A.4 Proof of Proposition 2.9

We first show the following equality

$$\hat{\mu}_\tau(n) + \sum_{s=1}^{n-1} \psi_{n-s}^\tau \hat{\mu}_\tau(s) = \mu_\tau + \xi \mu_\tau \frac{1}{1-a_\tau} + \mu_\tau (1-\xi) \sum_{s=1}^{n-1} \psi_{n-s}^\tau. \quad (\text{A.4})$$

In fact, using discrete convolution by φ^τ and fact that $\psi^\tau * \varphi^\tau = \psi^\tau - \varphi^\tau$, we obtain from the LHS of (A.4):

$$\begin{aligned} &\sum_{s=1}^{n-1} \hat{\mu}_\tau(s) \varphi_{n-s}^\tau + \sum_{s=1}^{n-1} \varphi_{n-s}^\tau \sum_{u=1}^{s-1} \psi_{s-u}^\tau \hat{\mu}_\tau(u) \\ &= \sum_{s=1}^{n-1} \hat{\mu}_\tau(s) \varphi_{n-s}^\tau + \sum_{u=1}^{n-1} \hat{\mu}_\tau(u) \sum_{s=1}^{n-u-1} \psi_s^\tau \varphi_{n-u-s}^\tau \\ &= \sum_{s=1}^{n-1} \hat{\mu}_\tau(s) \varphi_{n-s}^\tau + \sum_{u=1}^{n-1} (\psi_{n-u}^\tau - \varphi_{n-u}^\tau) \hat{\mu}_\tau(u) \\ &= \sum_{s=1}^{n-1} \psi_{n-s}^\tau \hat{\mu}_\tau(s). \end{aligned}$$

From RHS of (A.4), we get

$$\begin{aligned}
& \sum_{s=1}^{n-1} \varphi_{n-s}^\tau \left(\mu_\tau + \xi \mu_\tau \frac{1}{1-a_\tau} \right) + \mu_\tau (1-\xi) \sum_{s=1}^{n-1} \varphi_{n-s}^\tau \sum_{u=1}^{s-1} \psi_{s-u}^\tau \\
&= \mu_\tau \left(1 + \xi \frac{1}{1-a_\tau} \right) \sum_{s=1}^{n-1} \varphi_{n-s}^\tau + \mu_\tau (1-\xi) \sum_{u=1}^{n-1} \sum_{s=1}^{n-u-1} \psi_s^\tau \varphi_{n-u-s}^\tau \\
&= \mu_\tau \left(1 + \xi \frac{1}{1-a_\tau} \right) \sum_{s=1}^{n-1} \varphi_{n-s}^\tau + \mu_\tau (1-\xi) \sum_{u=1}^{n-1} (\psi_{n-u}^\tau - \varphi_{n-u}^\tau).
\end{aligned}$$

Consequently, we necessarily have

$$\sum_{s=1}^{n-1} \psi_{n-s}^\tau \hat{\mu}_\tau(s) = \mu_\tau \xi \left(\frac{1}{1-a_\tau} + 1 \right) \sum_{s=1}^{n-1} \varphi_{n-s}^\tau + \mu_\tau (1-\xi) \sum_{s=1}^{n-1} \psi_{n-s}^\tau.$$

The last equation together with (A.4) gives that

$$\hat{\mu}_\tau(n) = \mu_\tau + \xi \mu_\tau \frac{1}{1-a_\tau} \left(1 - \sum_{s=1}^{n-1} \varphi_{n-s}^\tau \right) - \mu_\tau \xi \sum_{s=1}^{n-1} \varphi_{n-s}^\tau.$$

Recall that $\lambda_n^{\tau,+} = \lambda_n^{\tau,-}$. By directly computing, we can write

$$\begin{aligned}
\lambda_n^{\tau,+} &= \mu_\tau + \mu_\tau \sum_{s=1}^{n-1} \psi_{n-s}^\tau + \xi \mu_\tau \left(\frac{1}{1-a_\tau} - \sum_{s=1}^{n-1} \psi_{n-s}^\tau \right) \\
&\quad + \frac{1}{\beta+1} \sum_{s=1}^{n-1} \psi_{n-s}^\tau (M_s^{\tau,+} - M_{s-1}^{\tau,+} + \beta M_s^{\tau,-} - \beta M_{s-1}^{\tau,-}).
\end{aligned}$$

Then using Tonelli's theorem and the fact that $\sum(\alpha * \beta) = (\sum \alpha) * \beta$, we get

$$\begin{aligned}
\sum_{s=1}^n \lambda_s^{\tau,+} &= \mu_\tau n + \mu_\tau \sum_{s=1}^{n-1} s \psi_{n-s}^\tau + \xi \mu_\tau \left(\frac{n}{1-a_\tau} - \sum_{s=1}^{n-1} s \psi_{n-s}^\tau \right) \\
&\quad + \frac{1}{\beta+1} \sum_{s=1}^{n-1} \psi_{n-s}^\tau (M_s^{\tau,+} + \beta M_s^{\tau,-}).
\end{aligned}$$

Therefore, for $t \in [0, 1]$, we have the decomposition

$$\Lambda_t^{\tau,+} = \Lambda_t^{\tau,-} = T_1 + T_2 + T_3, \tag{A.5}$$

with

$$\begin{aligned}
T_1 &:= (1-a_\tau)t, \\
T_2 &:= \frac{1-a_\tau}{\tau} \sum_{s=1}^{\lfloor \tau t \rfloor - 1} s \psi_{\lfloor \tau t \rfloor - s}^\tau + \xi \left(t - \frac{1-a_\tau}{\tau} \sum_{s=1}^{\lfloor \tau t \rfloor - 1} s \psi_{\lfloor \tau t \rfloor - s}^\tau \right), \\
T_3 &:= \frac{\tau(1-a_\tau)}{\sqrt{\gamma\mu(1+\beta)^2}} \sum_{s=1}^{\lfloor \tau t \rfloor - 1} \frac{1}{\tau} \psi_{\lfloor \tau t \rfloor - s}^\tau \left(Z_{s/\tau}^{\tau,+} + \beta Z_{s/\tau}^{\tau,-} \right).
\end{aligned}$$

Recall that

$$\frac{(1-a_\tau)}{a_\tau} \tau \psi_{\lfloor \tau \cdot \rfloor}^\tau \rightarrow f_{\alpha,\gamma} \text{ as } \tau \rightarrow \infty.$$

Thus, as $\tau \rightarrow \infty$,

$$T_2 \rightarrow \int_0^t f_{\alpha,\gamma}(t-s)sd s + \xi \left(t - \int_0^t f_{\alpha,\gamma}(t-s)sd s \right),$$

and

$$T_3 \rightarrow \frac{1}{\sqrt{\gamma\mu(1+\beta)^2}} \int_0^t f_{\alpha,\gamma}(t-s)(Z_s^+ + \beta Z_s^-)ds.$$

Therefore, letting $\tau \rightarrow \infty$, we obtain from Proposition 2.6 that Y satisfies

$$Y_t = \int_0^t f_{\alpha,\gamma}(t-s)sd s + \xi \left(t - \int_0^t f_{\alpha,\gamma}(t-s)sd s \right) + \frac{1}{\sqrt{\gamma\mu(1+\beta)^2}} \int_0^t f_{\alpha,\gamma}(t-s)(Z_s^+ + \beta Z_s^-)ds.$$

The rest of the proof follows similarly from the proof of Proposition 6.2 of [El Euch and Rosenbaum \(2019\)](#). This completes the proof of Proposition 2.9.

A.5 Proof of Theorem 2.10

The proof of Theorem 2.10 proceeds in two main steps, following the standard methodology for proving weak convergence of stochastic processes.

- *Step 1: Tightness.* In Proposition 2.6, we first establish that the sequence of processes $(\mathbf{\Lambda}^\tau, \mathbf{Y}^\tau, \mathbf{Z}^\tau)_{\tau \geq 1}$ is C -tight in the Skorokhod space. This ensures the existence of at least one convergent subsequence. The core of this step is Proposition 2.6, which also characterizes some fundamental properties of any possible limit point.
- *Step 2: Identification of the Limit.* In Lemma 2.8, we identify the limit of any such convergent subsequence. We show that any limit point must be the unique solution to the rough stochastic differential equation specified in the theorem statement. Since all subsequences converge to the same unique limit, this implies the convergence of the entire sequence.

With the tightness of the sequence $(\mathbf{Y}^\tau, \mathbf{Z}^\tau)$ and the characterization of its limit points established, we are now ready to conclude the proof of our main theorem.

Proposition 2.6 ensures that the sequence is tight and that any limit point (\mathbf{Y}, \mathbf{Z}) has the property that \mathbf{Z} is a continuous martingale with $[\mathbf{Z}, \mathbf{Z}] = \text{diag}(\mathbf{Y})$. Proposition 2.9 then uniquely identifies this limit point. It shows that the rate process v (where $Y_t = \int_0^t v_s ds$) must satisfy the stochastic Volterra equation (2.6).

Finally, we invoke Proposition 4.10 of [El Euch et al. \(2018\)](#), which proves that the solution to the stochastic Volterra equation (2.6) is equivalent to the solution of the rough differential equation given in the statement of Theorem 2.10.

Since the sequence is tight and all limit points are identical, the entire sequence converges in law to this unique limit. This completes the proof of Theorem 2.10.

A.6 Proof of Theorem 2.11

The proof follows the logic of Corollary 2.1 of [El Euch and Rosenbaum \(2019\)](#), leveraging the convergence result established in our Theorem 2.10 and applying the continuous mapping theorem. We outline the key steps.

First, recall the definition of the properly rescaled microscopic price process P_t^τ :

$$\begin{aligned} \sqrt{\frac{1-a_\tau}{\mu\tau^\alpha}} P_t^\tau &= \sqrt{\frac{\theta}{2}} \sqrt{\frac{1-a_\tau}{\tau^\alpha \mu}} (N_t^{\tau,+} - N_t^{\tau,-}) - \frac{\theta}{2} \frac{1-a_\tau}{\tau^\alpha \mu} N_t^{\tau,+} \\ &= \sqrt{\frac{\theta}{2}} (Y_t^{\tau,+} - \Lambda_t^{\tau,+} - (Y_t^{\tau,-} - \Lambda_t^{\tau,-})) - \frac{\theta}{2} Y_t^{\tau,+} \end{aligned}$$

$$= \frac{\sqrt{\theta}}{\sqrt{\frac{\tau^\alpha \mu}{1-a_\tau}}} \left(\frac{\sqrt{\theta}}{2} (Z_t^{\tau,+} - Z_t^{\tau,-}) - \frac{\theta}{2} Y_t^{\tau,+} \right).$$

Rearranging the terms, we have

$$P_t^\tau = \frac{\theta}{2} (Z_t^{\tau,+} - Z_t^{\tau,-}) - \frac{\theta}{2} Y_t^{\tau,+}.$$

The mapping from the process triplet $(Z_t^{\tau,+}, Z_t^{\tau,-}, Y_t^{\tau,+})$ to P_t^τ is a continuous linear combination. Theorem 2.10 establishes the joint convergence in law of these processes to their limits (Z_t^+, Z_t^-, Y_t^+) in the Skorokhod topology. By the continuous mapping theorem, P_t^τ converges in law to a process P_t given by the same linear combination of the limits:

$$\begin{aligned} P_t &= \frac{\theta}{2} (Z_t^+ - Z_t^-) - \frac{\theta}{2} Y_t^+ \\ &= \frac{\theta}{2} \left(\int_0^t \sqrt{v_s} dB_s^1 - \int_0^t \sqrt{v_s} dB_s^2 \right) - \frac{\theta}{2} \int_0^t v_s ds \\ &= \sqrt{\frac{\theta}{2}} \int_0^t \sqrt{\theta v_s} \frac{1}{\sqrt{2}} (dB_s^1 - dB_s^2) - \frac{1}{2} \int_0^t (\theta v_s) ds, \end{aligned}$$

where we have substituted the integral representations of the limit processes from Theorem 2.10.

Now, we define the variance process $V_t := \theta v_t$ and a new Brownian motion $W_t^X := \frac{1}{\sqrt{2}}(B_t^1 - B_t^2)$. With these definitions, the expression for P_t becomes:

$$P_t = \int_0^t \sqrt{V_s} dW_s^X - \frac{1}{2} \int_0^t V_s ds.$$

This is precisely the log-price process in the rough Heston model under the risk-neutral measure.

To obtain the SDE for the variance process V_t , we simply multiply the SDE for v_t from Theorem 2.10 by the constant θ . This yields equation (2.13).

Finally, we compute the correlation between the Brownian motions W^X and W^V . Using the properties of quadratic covariation and the definitions of W_t^X and $W_t^V = (B_t^1 + \beta B_t^2)/\sqrt{1 + \beta^2}$, we have:

$$\begin{aligned} d\langle W^X, W^V \rangle_t &= d\left\langle \frac{1}{\sqrt{2}} (B^1 - B^2), \frac{1}{\sqrt{1 + \beta^2}} (B^1 + \beta B^2) \right\rangle_t \\ &= \frac{1}{\sqrt{2(1 + \beta^2)}} (d\langle B^1, B^1 \rangle_t + \beta d\langle B^1, B^2 \rangle_t - d\langle B^2, B^1 \rangle_t - \beta d\langle B^2, B^2 \rangle_t). \end{aligned}$$

Since B^1 and B^2 are independent standard Brownian motions, we have $d\langle B^1, B^1 \rangle_t = d\langle B^2, B^2 \rangle_t = dt$ and $d\langle B^1, B^2 \rangle_t = 0$. This simplifies to:

$$d\langle W^X, W^V \rangle_t = \frac{1 - \beta}{\sqrt{2(1 + \beta^2)}} dt.$$

This completes the proof.

B Proof of Section 3

B.1 Proof of Lemma 3.1

Write $a_n := \frac{1}{\Gamma(1-\alpha)} \frac{1}{n^\alpha}$ and note

$$\sum_{n \geq 2} \varphi_n e^{-sn} = \sum_{n \geq 2} (a_{n-1} - a_n) e^{-sn} = e^{-s} \sum_{m \geq 1} a_m e^{-sm} - \sum_{n \geq 2} a_n e^{-sn} = -(1 - e^{-s}) \sum_{m \geq 1} a_m e^{-sm} + a_1 e^{-s}.$$

Adding $\varphi_1 e^{-s} = (1 - a_1)e^{-s}$ gives the exact identity

$$\hat{\varphi}(s) = e^{-s} - (1 - e^{-s})A(s), \quad A(s) := \sum_{m \geq 1} a_m e^{-sm} = \frac{1}{\Gamma(1 - \alpha)} \text{Li}_\alpha(e^{-s}),$$

where Li_α is the polylogarithm. For $0 < \alpha < 1$ one has the classical expansion (see, e.g., [Olver et al. \(2023\)](#)):

$$\text{Li}_\alpha(e^{-s}) = \Gamma(1 - \alpha)s^{\alpha-1} + \zeta(\alpha) + \mathcal{O}(s) \quad (s \downarrow 0).$$

Hence

$$A(s) = s^{\alpha-1} + \frac{\zeta(\alpha)}{\Gamma(1 - \alpha)} + \mathcal{O}(s), \quad 1 - e^{-s} = s - \frac{s^2}{2} + \mathcal{O}(s^3), \quad e^{-s} = 1 - s + \frac{s^2}{2} + \mathcal{O}(s^3).$$

Plugging into $\hat{\varphi}(s) = e^{-s} - (1 - e^{-s})A(s)$ yields

$$\hat{\varphi}(s) = 1 - s^\alpha - c_1 s + \mathcal{O}(s^{\alpha+1} + s^2), \quad c_1 := 1 + \frac{\zeta(\alpha)}{\Gamma(1 - \alpha)}.$$

Therefore $R(s) = -c_1 s + \mathcal{O}(s^{\alpha+1} + s^2)$ so that $|R(s)| \leq C_R s$ for small s . Since $\alpha + \delta \leq 1$, we also have $s \leq s^{\alpha+\delta}$ for small s ; hence $|R(s)| \leq C_R s^{\alpha+\delta}$ for any $\delta \in (0, 1 - \alpha]$. This completes the proof.

B.2 Proof of Lemma 3.3

Step 1: A small-parameter exponential moment for $\int_0^T V_s ds$. For affine-Volterra square-root models one has the mild form and, by stochastic Fubini's theorem (see [Veraar \(2012\)](#) and [\(Abi Jaber et al., 2019, Lem. 2.6\)](#)),

$$\int_0^T V_s ds = D_T + c_V \int_0^T \Phi(s) \sqrt{V_s} dW_s^V,$$

where D_T is a deterministic finite constant and $\Phi(s) = \int_0^{T-s} f_{\alpha,\gamma}(u) du$. Let $M(\lambda) := \mathbb{E} \left[\exp \left(\lambda \int_0^T V_s ds \right) \right]$ for $\lambda \geq 0$. Define the martingale $M_t := \lambda c_V \int_0^t \Phi(s) \sqrt{V_s} dW_s^V$, so that $\int_0^T V_s ds = D_T + \frac{1}{\lambda} M_T$. The quadratic variation is $\langle M \rangle_T = \lambda^2 c_V^2 \int_0^T \Phi^2(s) V_s ds$. We can write

$$M(\lambda) = e^{\lambda D_T} \mathbb{E} [\exp(M_T)].$$

To bound $\mathbb{E}[\exp(M_T)]$, we use the supermartingale property of stochastic exponentials. We consider the decomposition $\exp(M_T) = A \cdot B$, where:

$$A := \exp(M_T - \langle M \rangle_T), \quad B := \exp(\langle M \rangle_T).$$

By the Cauchy-Schwarz inequality, $\mathbb{E}[\exp(M_T)] = \mathbb{E}[A \cdot B] \leq (\mathbb{E}[A^2])^{1/2} (\mathbb{E}[B^2])^{1/2}$.

We analyze the $\mathbb{E}[A^2]$ term first. Let $N_T := 2M_T$. Its quadratic variation is $\langle N \rangle_T = 4\langle M \rangle_T$.

$$A^2 = \exp(2M_T - 2\langle M \rangle_T) = \exp \left(N_T - \frac{1}{2} \langle N \rangle_T \right) = \mathcal{E}(N)_T.$$

Since A^2 is itself a positive stochastic exponential (a supermartingale), we have

$$\mathbb{E}[A^2] = \mathbb{E}[\mathcal{E}(N)_T] \leq 1.$$

Next, we analyze the $\mathbb{E}[B^2]$ term.

$$\mathbb{E}[B^2] = \mathbb{E}[\exp(2\langle M \rangle_T)]$$

$$\begin{aligned}
&= \mathbb{E} \left[\exp \left(2\lambda^2 c_V^2 \int_0^T (\Phi(s))^2 V_s ds \right) \right] \\
&\leq \mathbb{E} \left[\exp \left(2\lambda^2 c_V^2 \|\Phi\|_{L^\infty(0,T)}^2 \int_0^T V_s ds \right) \right].
\end{aligned}$$

Let $c := \frac{1}{2} c_V^2 \|\Phi\|_{L^\infty(0,T)}^2$. Then $2\lambda^2 c_V^2 \|\Phi\|_{L^\infty(0,T)}^2 = 4c\lambda^2$. Thus, $\mathbb{E}[B^2] \leq \mathbb{E} \left[\exp \left(4c\lambda^2 \int_0^T V_s ds \right) \right] = M(4c\lambda^2)$. Combining these results, we get

$$\mathbb{E}[\exp(M_T)] \leq (1)^{1/2} (M(4c\lambda^2))^{1/2} = \sqrt{M(4c\lambda^2)}.$$

Substituting this back into the expression for $M(\lambda)$:

$$M(\lambda) \leq e^{\lambda D_T} \sqrt{M(4c\lambda^2)}, \quad \text{or} \quad (M(\lambda))^2 \leq e^{2\lambda D_T} M(4c\lambda^2).$$

This recursion shows that $M(\lambda)$ is finite. If we start with $\lambda_0 = \lambda$ and define the sequence $\lambda_{n+1} = 4c\lambda_n^2$, then $M(\lambda_0)$ is finite as long as $\lambda_n \rightarrow 0$. This holds if $4c\lambda_0 < 1$, or $\lambda_0 < 1/(4c)$.

Thus, $M(\lambda) < \infty$ for all $\lambda \in [0, \Theta_T/4]$.

Step 2: A vertical strip around $\text{Re}(z) = \frac{1}{2}$ via exponential martingale bounds. Write the log-price dynamics as

$$dX_t = \left(-\frac{1}{2} c_X^2 V_t \right) dt + c_X \sqrt{V_t} dW_t^X.$$

Fix $z \in \mathbb{C}$ and set $y := \text{Re}(z)$. Since $|\mathbb{E}[e^{zX_T}]| \leq \mathbb{E}[e^{yX_T}]$, it suffices to bound $\mathbb{E}[e^{yX_T}]$ for real y . We can compute that

$$\mathbb{E}[e^{yX_T}] = e^{yX_0} \mathbb{E} \left[\exp \left(y c_X \int_0^T \sqrt{V_s} dW_s^X - \frac{1}{2} y c_X^2 \int_0^T V_s ds \right) \right].$$

We decompose the term inside the expectation as $A \cdot B$, that is,

$$\mathbb{E}[e^{yX_T}] = e^{yX_0} \cdot \mathbb{E}[A \cdot B],$$

where

$$\begin{aligned}
A &:= \exp \left(y c_X \int_0^T \sqrt{V_s} dW_s^X - y^2 c_X^2 \int_0^T V_s ds \right), \\
B &:= \exp \left(\left(y^2 - \frac{1}{2} y \right) c_X^2 \int_0^T V_s ds \right),
\end{aligned}$$

and apply the Cauchy-Schwarz inequality $\mathbb{E}[A \cdot B] \leq (\mathbb{E}[A^2])^{1/2} (\mathbb{E}[B^2])^{1/2}$.

First, we analyze $\mathbb{E}[A^2]$. Let $M_t := 2y c_X \int_0^t \sqrt{V_s} dW_s^X$. Its quadratic variation is $\langle M \rangle_T = 4y^2 c_X^2 \int_0^T V_s ds$.

$$A^2 = \exp \left(2y c_X \int_0^T \sqrt{V_s} dW_s^X - 2y^2 c_X^2 \int_0^T V_s ds \right) = \exp \left(M_T - \frac{1}{2} \langle M \rangle_T \right) = \mathcal{E}(M)_T.$$

Since A^2 is exactly the stochastic exponential $\mathcal{E}(M)_T$, which is a positive supermartingale, we have $\mathbb{E}[A^2] \leq 1$.

Now, by applying the Cauchy-Schwarz inequality and using $\mathbb{E}[A^2] \leq 1$, we obtain:

$$\begin{aligned}
\mathbb{E}[e^{yX_T}] &\leq e^{yX_0} (\mathbb{E}[A^2])^{1/2} (\mathbb{E}[B^2])^{1/2} \\
&\leq e^{yX_0} (1)^{1/2} \left\{ \mathbb{E} \left[\exp \left(2 \left(y^2 - \frac{1}{2} y \right) c_X^2 \int_0^T V_s ds \right) \right] \right\}^{1/2}
\end{aligned}$$

$$= e^{yX_0} \left(\mathbb{E} \exp \left((2y^2 - y) c_X^2 \int_0^T V_s ds \right) \right)^{1/2}.$$

Let $M(\lambda) = \mathbb{E}[\exp(\lambda \int_0^T V_s ds)]$ be the moment generating function from Step 1. We have $\mathbb{E}[e^{yX_T}] \leq e^{yX_0} \sqrt{M(\lambda_y)}$, where $\lambda_y := (2y^2 - y)c_X^2$. From Step 1, we know $M(\lambda)$ is finite for $\lambda \in [0, \Theta_{\text{eff}})$, where $\Theta_{\text{eff}} = \Theta_T/4 > 0$. Therefore, $\mathbb{E}[e^{yX_T}]$ is finite as long as $\lambda_y < \Theta_T/4$. That is,

$$(2y^2 - y)c_X^2 < \Theta_T/4.$$

We need to show this condition holds for y in an open interval containing $y = 1/2$. Let $f(y) = (2y^2 - y)c_X^2$. At $y = 1/2$, we have

$$f(1/2) = (2(1/4) - 1/2)c_X^2 = 0.$$

Since $f(1/2) = 0$ and $0 < \Theta_T/4$ (as $\Theta_T > 0$), by the continuity of $f(y)$, there must exist an open interval $(\frac{1}{2} - \delta, \frac{1}{2} + \delta)$ for some $\delta > 0$ such that $f(y) < \Theta_T/4$ for all y in this interval.

This establishes that $\mathbb{E}[e^{yX_T}] < \infty$ for all $y \in (\frac{1}{2} - \delta, \frac{1}{2} + \delta)$. Since $|\mathbb{E}[e^{zX_T}]| \leq \mathbb{E}[e^{yX_T}]$ with $y = \text{Re}(z)$, the map $z \mapsto \mathbb{E}[e^{zX_T}]$ is finite and locally dominated on this open strip, hence analytic there.

Step 3: Affine-Volterra transform and Riccati well-posedness. By the affine-Volterra transform formula (Abi Jaber et al., 2019, Thm. 4.3) (see also (Abi Jaber et al., 2019, Thm. 6.1) for the Volterra square-root setting), for each z in the open strip $\{\frac{1}{2} - \delta < \text{Re}(z) < \frac{1}{2} + \delta\}$ identified in Step 2 (where $\delta > 0$), the moment generating function $\mathbb{E}[e^{zX_T}]$ is analytic. The transform theorem thus implies that there exist unique functions $(\mathcal{A}(\cdot, z), \mathcal{B}(\cdot, z))$ solving the affine-Volterra Riccati system on $[0, T]$ with

$$\mathbb{E}[e^{zX_T}] = \exp(\mathcal{B}(T, z) + V_0 \mathcal{A}(T, z)).$$

The analyticity of the map $z \mapsto \mathcal{B}(T, z) + V_0 \mathcal{A}(T, z)$ on the open strip follows from that of the MGF. The uniform bound on its real part, as stated in (ii) of the lemma, follows directly from the estimate $\mathbb{E}[e^{yX_T}] \leq e^{yX_0} \sqrt{M(\lambda_y)} < \infty$ derived in Step 2. This proves (i) and (ii).

B.3 Proof of Proposition 3.5

Write $F^\tau(t) := \int_0^t \zeta^\tau(u) du$ and $\Phi^\tau(s) := F^\tau(T - s) \geq 0$. By the uniform L^2 -energy bound (3.8), we know

$$\sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0, T)} \leq \sup_{\tau \geq 1} \int_0^T \zeta^\tau(u) du \leq \sqrt{T} \left(\sup_{\tau \geq 1} \int_0^T (\zeta^\tau(u))^2 du \right)^{1/2} < \infty.$$

Step 1: Uniform small-parameter exponential moment for $\int_0^T V_s^{(\tau)} ds$. As in the proof of Lemma 3.3, stochastic Fubini's theorem yields

$$\int_0^T V_s^{(\tau)} ds = D_T + c_V \int_0^T \Phi^\tau(s) \sqrt{V_s^{(\tau)}} dW_s^V,$$

with a deterministic D_T . For $\lambda \geq 0$, let $M_\tau(\lambda) := \mathbb{E} \exp \left(\lambda \int_0^T V_s^{(\tau)} ds \right)$. Let $M_t := \lambda c_V \int_0^t \Phi^\tau(s) \sqrt{V_s^{(\tau)}} dW_s^V$ and $\langle M \rangle_T = \lambda^2 c_V^2 \int_0^T (\Phi^\tau(s))^2 V_s^{(\tau)} ds$. We have $M_\tau(\lambda) = e^{\lambda D_T} \mathbb{E}[\exp(M_T)]$. We apply the Cauchy-Schwarz inequality to $\mathbb{E}[\exp(M_T)] = \mathbb{E}[A \cdot B]$, with

$$A := \exp(M_T - \langle M \rangle_T), \quad B := \exp(\langle M \rangle_T).$$

As shown in the proof of Lemma 3.3, $\mathbb{E}[A^2] = \mathbb{E}[\mathcal{E}(2M)_T] \leq 1$. For the $\mathbb{E}[B^2]$ term, we have

$$\mathbb{E}[B^2] = \mathbb{E}[\exp(2\langle M \rangle_T)] = \mathbb{E} \left[\exp \left(2\lambda^2 c_V^2 \int_0^T (\Phi^\tau(s))^2 V_s^{(\tau)} ds \right) \right]$$

$$\leq \mathbb{E} \left[\exp \left(2\lambda^2 c_V^2 \|\Phi^\tau\|_{L^\infty(0,T)}^2 \int_0^T V_s^{(\tau)} ds \right) \right].$$

Let $c^* := \frac{1}{2} c_V^2 \sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0,T)}^2$. By assumption (3.8), $c^* < \infty$. Then $\mathbb{E}[B^2] \leq M_\tau(4c^*\lambda^2)$. This gives the uniform recursion

$$M_\tau(\lambda) \leq e^{\lambda D_\tau} (M_\tau(4c^*\lambda^2))^{1/2}, \quad \text{or equivalently} \quad (M_\tau(\lambda))^2 \leq e^{2\lambda D_\tau} M_\tau(4c^*\lambda^2).$$

Define the uniform threshold $\Theta^* := \frac{1}{c^*} = \frac{2}{c_V^2 \sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0,T)}^2} > 0$. The recursion shows that $M_\tau(\lambda)$ is finite (with a bound uniform in τ) for all $\lambda \in [0, \Theta_{\text{eff}}^*]$, where $\Theta_{\text{eff}}^* = 1/(4c^*) = \Theta^*/4$.

Step 2: Uniform vertical strip around $\text{Re}(z) = \frac{1}{2}$. Write

$$dX_t^{(\tau)} = -\frac{1}{2} c_X^2 V_t^{(\tau)} dt + c_X \sqrt{V_t^{(\tau)}} dW_t^X,$$

and fix $z \in \mathbb{C}$, $y := \text{Re}(z)$. By following the argument as in the proof of Lemma 3.3, we write

$$\mathbb{E} \left[e^{yX_T^{(\tau)}} \right] = e^{yX_0} \mathbb{E}[A \cdot B], \tag{B.1}$$

where

$$\begin{aligned} A &:= \exp \left(y c_X \int_0^T \sqrt{V_s^{(\tau)}} dW_s^X - y^2 c_X^2 \int_0^T V_s^{(\tau)} ds \right), \\ B &:= \exp \left(\left(y^2 - \frac{1}{2} y \right) c_X^2 \int_0^T V_s^{(\tau)} ds \right). \end{aligned}$$

Note that we have $\mathbb{E}[A^2] \leq 1$ (uniformly in τ). By applying Cauchy-Schwarz inequality, this yields the bound

$$\mathbb{E} \left[e^{yX_T^{(\tau)}} \right] \leq e^{yX_0} (\mathbb{E}[A^2])^{1/2} (\mathbb{E}[B^2])^{1/2} \leq e^{yX_0} \left(\mathbb{E} \exp \left((2y^2 - y) c_X^2 \int_0^T V_s^{(\tau)} ds \right) \right)^{1/2}.$$

This is bounded (uniformly in τ) by $e^{yX_0} \sqrt{M_\tau(\lambda_y)}$, where $\lambda_y := (2y^2 - y)c_X^2$. From Step 1, this is finite (uniformly in τ) as long as $\lambda_y < \Theta_{\text{eff}}^* = \Theta^*/4$. The condition is $(2y^2 - y)c_X^2 < \Theta^*/4$. Since this condition is satisfied for $y = 1/2$ (giving $0 < \Theta^*/4$), it holds for a uniform open interval $(\frac{1}{2} - \delta, \frac{1}{2} + \delta)$, where

$$\delta = \frac{\sqrt{1 + 2\Theta^*/c_X^2} - 1}{4} > 0.$$

Since Θ^* is independent of τ , this δ is also independent of τ .

Step 3: Affine-Volterra transform and Riccati well-posedness (uniform in τ). Applying the affine-Volterra transform (Abi Jaber et al., 2019, Thm. 4.3) for each kernel ζ^τ , we obtain

$$\mathbb{E} \left[e^{zX_T^{(\tau)}} \right] = \exp(\mathcal{B}^\tau(T, z) + V_0 \mathcal{A}^\tau(T, z)).$$

By the uniform bound from Step 2, there exists $\delta > 0$ and $C_T < \infty$, both independent of τ , such that

$$\sup_{\tau \geq 1} \sup_{\text{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]} \text{Re}(\mathcal{B}^\tau(T, z) + V_0 \mathcal{A}^\tau(T, z)) \leq C_T.$$

This implies the conclusions of the proposition hold.

B.4 Proof of Proposition 3.8

Let $g(t) := \zeta^\tau(t) - f_{\alpha,\gamma}(t)$ and fix $\sigma > 0$. Set $z := \sigma + i\omega$ and $s := z/\tau$. Recall from the proof of Lemma 2.5 (see Appendix A.1)

$$\hat{\zeta}^\tau(z) = (1 - a_\tau) \frac{\tau(1 - e^{-z/\tau})}{z} \frac{\hat{\varphi}(s)}{1 - a_\tau \hat{\varphi}(s)}, \quad \hat{f}_{\alpha,\gamma}(z) = \frac{\gamma}{\gamma + z^\alpha}.$$

Write $G(z) := \frac{\tau(1 - e^{-z/\tau})}{z}$; then

$$|G(z)| \leq \min\left(1, \frac{\tau}{|z|}\right), \quad |G(z) - 1| \leq \min\left(\frac{|z|}{\tau}, 1\right).$$

Using Lemma 3.1, $\hat{\varphi}(s) = 1 - s^\alpha + R(s)$ with $|R(s)| \leq C_R |s|^{\alpha+\delta}$ as $s \downarrow 0$; moreover

$$1 - a_\tau \hat{\varphi}(s) = (1 - a_\tau) + a_\tau s^\alpha - a_\tau R(s) = \tau^{-\alpha} (\gamma + a_\tau z^\alpha) - a_\tau R(s).$$

Hence for τ large enough,

$$|1 - a_\tau \hat{\varphi}(s)| \gtrsim \tau^{-\alpha} (1 + |z|^\alpha).$$

A direct algebraic decomposition gives

$$\hat{\zeta}^\tau(z) - \hat{f}_{\alpha,\gamma}(z) = \frac{\gamma [(G - 1)\gamma + (G - 1)z^\alpha + (1 - a_\tau)z^\alpha - G(\gamma + z^\alpha) \frac{z^\alpha}{\tau^\alpha} + G(\gamma + z^\alpha)R(s) + \tau^\alpha a_\tau R(s)]}{(\gamma + z^\alpha)(\gamma + a_\tau z^\alpha - \tau^\alpha a_\tau R(s))}.$$

Using the bounds above together with $|R(s)| \leq C_R |z|^{\alpha+\delta} \tau^{-(\alpha+\delta)}$ and $|(\gamma + z^\alpha)| \gtrsim 1 + |z|^\alpha$, we obtain the piecewise estimate

$$\left| \hat{\zeta}^\tau(\sigma + i\omega) - \hat{f}_{\alpha,\gamma}(\sigma + i\omega) \right| \leq \begin{cases} C_0 \tau^{-\alpha}, & |\omega| \leq \tau, \\ C_0 \tau^{1-\alpha} |\omega|^{-1}, & |\omega| > \tau, \end{cases} \quad (\text{B.2})$$

for a constant C_0 depending only on $(\alpha, \gamma, \delta, C_R)$. (The term $-G(\gamma + z^\alpha)z^\alpha/\tau^\alpha$ yields the worst high-frequency behavior $\sim \tau^{1-\alpha}/|\omega|$, while the remaining terms are of order $\tau^{-\alpha}$ or higher order in τ^{-1} after division by the denominator.)

By the Paley–Wiener–Plancherel identity,

$$\|e^{-\sigma \cdot} g\|_{L^2(0,\infty)}^2 = \frac{1}{2\pi} \int_{\mathbb{R}} \left| \hat{\zeta}^\tau(\sigma + i\omega) - \hat{f}_{\alpha,\gamma}(\sigma + i\omega) \right|^2 d\omega.$$

Using (B.2) and splitting this integral at $|\omega| = \tau$

$$\int_{|\omega| \leq \tau} \tau^{-2\alpha} d\omega + \int_{|\omega| > \tau} \tau^{2-2\alpha} \omega^{-2} d\omega \lesssim \tau^{1-2\alpha}.$$

Therefore $\|e^{-\sigma \cdot} g\|_{L^2(0,\infty)} \lesssim \tau^{\frac{1}{2}-\alpha}$, and hence

$$\|g\|_{L^2(0,1)} = \left(\int_0^1 |g(t)|^2 dt \right)^{1/2} \leq e^\sigma \left(\int_0^1 e^{-2\sigma t} |g(t)|^2 dt \right)^{1/2} \leq e^\sigma \|e^{-\sigma \cdot} g\|_{L^2(0,\infty)} \leq C_1 \tau^{\frac{1}{2}-\alpha}.$$

Finally, by Cauchy–Schwarz inequality,

$$\sup_{t \leq 1} \int_0^t |g(u)| du \leq \|g\|_{L^2(0,1)} \leq C_2 \tau^{\frac{1}{2}-\alpha}.$$

B.5 Proof of Proposition 3.9

Bracket bias. In the compound-Poisson layer with intensity $\lambda_\tau = \tau\lambda$ and centered jumps of variance σ_J^2 with $\lambda\sigma_J^2 = 1$, we have $[M^\tau]_t = \lfloor t\tau \rfloor / \tau$, hence $\sup_{t \leq T} |\mathbb{E}[M^\tau]_t - t| \leq \tau^{-1}$. In the near-unstable INAR(1) layer (thinning $1 - 1/\tau$ with i.i.d. innovations of finite fourth moment), a one-step calculation yields the conditional variance $v_k^\tau = 1 + \mathcal{O}(\tau^{-1})$ uniformly in k on $[0, T]$, so that $\sup_{t \leq T} |\mathbb{E}[M^\tau]_t - t| \leq C_T \tau^{-1}$.

Time-averaged L^2 bound. Realize M^τ as a martingale difference array $M_t^\tau = \tau^{-1/2} \sum_{k \leq \lfloor t\tau \rfloor} \xi_k^\tau$ with $\mathbb{E}[\xi_k^\tau | \mathcal{F}_{k-1}^\tau] = 0$ and $[M^\tau]_t = \tau^{-1} \sum_{k \leq \lfloor t\tau \rfloor} v_k^\tau$, $v_k^\tau := \mathbb{E}[(\xi_k^\tau)^2 | \mathcal{F}_{k-1}^\tau]$. Apply a Skorokhod embedding (see e.g. Hall and Heyde (1980)) on an enlarged space to construct a Brownian motion W and stopping times S_k^τ with $S_0^\tau = 0$ such that

$$W_{S_k^\tau} - W_{S_{k-1}^\tau} \stackrel{d}{=} \frac{1}{\sqrt{\tau}} \xi_k^\tau, \quad \mathbb{E}[S_k^\tau - S_{k-1}^\tau | \mathcal{F}_{k-1}^\tau] = \frac{1}{\tau} v_k^\tau,$$

and a second-moment control

$$\mathbb{E} \left[\left(S_k^\tau - S_{k-1}^\tau - \frac{1}{\tau} v_k^\tau \right)^2 \right] \leq C \tau^{-2},$$

uniformly in k (using the finite fourth moments of ξ_k^τ). Let $A_t^\tau := S_{\lfloor t\tau \rfloor}^\tau$. Then

$$\mathbb{E}[|M_t^\tau - W_t|^2] = \mathbb{E}[|W_{A_t^\tau} - W_t|^2] = \mathbb{E}[|A_t^\tau - t|].$$

By the triangle inequality and Cauchy–Schwarz inequality,

$$\mathbb{E}[|A_t^\tau - t|] \leq \left(\mathbb{E}[|A_t^\tau - [M^\tau]_t|^2] \right)^{1/2} + |\mathbb{E}[M^\tau]_t - t|.$$

Since $\mathbb{E}[A_t^\tau - [M^\tau]_t] = 0$ and

$$A_t^\tau - [M^\tau]_t = \sum_{k \leq \lfloor t\tau \rfloor} \left(S_k^\tau - S_{k-1}^\tau - \frac{1}{\tau} v_k^\tau \right),$$

the orthogonality of martingale differences yields

$$\mathbb{E}[|A_t^\tau - [M^\tau]_t|^2] = \sum_{k \leq \lfloor t\tau \rfloor} \mathbb{E} \left[\left(S_k^\tau - S_{k-1}^\tau - \frac{1}{\tau} v_k^\tau \right)^2 \right] \leq C t \tau^{-1}.$$

Hence

$$\left(\mathbb{E}[|A_t^\tau - [M^\tau]_t|^2] \right)^{1/2} \leq C \tau^{-1/2}.$$

Combining with the bracket bias bound gives

$$\mathbb{E}[|M_t^\tau - W_t|^2] = \mathbb{E}[|W_{A_t^\tau} - W_t|^2] = \mathbb{E}[|A_t^\tau - t|] \leq C \tau^{-1/2} + C_T \tau^{-1}.$$

Integrating over $t \in [0, T]$ yields

$$\int_0^T \mathbb{E}[|M_t^\tau - W_t|^2] dt \leq C_T \tau^{-1/2},$$

which proves the stated rate.

B.6 Proof of Theorem 3.10

Step 1: Driver-to-state stability. Let $\Delta V_t := \widehat{V}_t^\tau - V_t^\tau$, $\Delta X_t := \widehat{X}_t^\tau - X_t^\tau$, and $\Delta M_t := (M_t^{\tau,X} - W_t^X, M_t^{\tau,V} - W_t^V)$. Using the mild Volterra forms and the same coefficients for the two systems (which are the ζ^τ -analogues of (3.5) and (3.6)),

$$\Delta V_t = c_V \int_0^t \zeta^\tau(t-s) \left(\sqrt{\widehat{V}_s^\tau} - \sqrt{V_s^\tau} \right) dW_s^V + c_V \int_0^t \zeta^\tau(t-s) \sqrt{\widehat{V}_s^\tau} d\Delta M_s^V. \quad (\text{B.3})$$

By $(\sqrt{a} - \sqrt{b})^2 \leq |a - b|$ for any $a, b \geq 0$ and Burkholder-Davis-Gundy (BDG) inequality in L^2 ,

$$\mathbb{E}|\Delta V_t| \leq C \left(\int_0^t \zeta^\tau(t-s)^2 \mathbb{E}|\Delta V_s| ds \right)^{1/2} + C \left(\int_0^t \zeta^\tau(t-s)^2 \mathbb{E} \left[\widehat{V}_s^\tau \right] \mathbb{E} |\Delta M_s^V|^2 ds \right)^{1/2}.$$

First, we claim that uniform moment bounds for \widehat{V}^τ on $[0, T]$ (from the moment strip) and $\int_0^T (\zeta^\tau(s))^2 ds \leq C_T$ yield, via Young's inequality and a Volterra-Grönwall lemma (Lemma 3.2),

$$\sup_{t \leq T} \mathbb{E}|\Delta V_t| \leq C_T \left(\int_0^T \mathbb{E} |\Delta M_s^V|^2 ds \right)^{1/2}. \quad (\text{B.4})$$

Let us prove (B.4). By adding and subtracting the term $c_V \int_0^t \zeta^\tau(t-s) \sqrt{\widehat{V}_s^\tau} dW_s^V$ in (B.3), we get the decomposition

$$\begin{aligned} \Delta V_t = & \underbrace{\int_0^t \zeta^\tau(t-s) \left(b(\widehat{V}_s^\tau) - b(V_s^\tau) \right) ds}_{=: I_1(t)} + \underbrace{c_V \int_0^t \zeta^\tau(t-s) \left(\sqrt{\widehat{V}_s^\tau} - \sqrt{V_s^\tau} \right) dW_s^V}_{=: I_2(t)} \\ & + \underbrace{c_V \int_0^t \zeta^\tau(t-s) \sqrt{\widehat{V}_s^\tau} d(M_s^{\tau,V} - W_s^V)}_{=: I_3(t)}. \end{aligned}$$

Let $u(t) := \sup_{r \leq t} \mathbb{E}|\Delta V_r|$. Using the uniform moment bounds for \widehat{V}^τ, V^τ on $[0, T]$ (from the moment strip) and $\int_0^T (\zeta^\tau(s))^2 ds \leq C_T$, we estimate the three terms as follows.

If b is Lipschitz on \mathbb{R}_+ with constant L_b , then

$$\mathbb{E}|I_1(t)| \leq L_b \int_0^t |\zeta^\tau(t-s)| \mathbb{E}|\Delta V_s| ds \leq L_b \|\zeta^\tau\|_{L^1(0,T)} u(t) \leq C_T u(t),$$

where we used Cauchy-Schwarz inequality to get $\|\zeta^\tau\|_{L^1(0,T)} \leq \sqrt{T} \|\zeta^\tau\|_{L^2(0,T)}$.

By BDG inequality and $(\sqrt{a} - \sqrt{b})^2 \leq |a - b|$ for any $a, b \geq 0$,

$$\mathbb{E}|I_2(t)| \leq C \mathbb{E} \left[\left(\int_0^t (\zeta^\tau(t-s))^2 \left(\sqrt{\widehat{V}_s^\tau} - \sqrt{V_s^\tau} \right)^2 ds \right)^{1/2} \right] \leq C \left(\int_0^t (\zeta^\tau(s))^2 ds \right)^{1/2} \left(\int_0^t \mathbb{E}|\Delta V_s| ds \right)^{1/2}.$$

Hence, for any $\varepsilon \in (0, 1)$, Young's inequality $ab \leq \varepsilon a^2 + \frac{1}{4\varepsilon} b^2$ yields

$$\mathbb{E}|I_2(t)| \leq \varepsilon u(t) + C_{T,\varepsilon} \int_0^t (\zeta^\tau(s))^2 ds \leq \varepsilon u(t) + C_{T,\varepsilon}.$$

By BDG inequality and Kunita-Watanabe inequality,

$$\mathbb{E}|I_3(t)| \leq C \left(\mathbb{E} \int_0^t (\zeta^\tau(t-s))^2 \widehat{V}_s^\tau d \langle M^{\tau,V} - W^V \rangle_s \right)^{1/2} \leq C \left(\int_0^t (\zeta^\tau(s))^2 ds \right)^{1/2} \left(\int_0^t \mathbb{E} |\Delta M_s^V|^2 ds \right)^{1/2},$$

where we also used the uniform bound $\sup_{s \leq T} \mathbb{E} \widehat{V}_s^\tau \leq C_T$ and the time-averaged L^2 control on ΔM^V .

Collecting the three estimates and choosing $\varepsilon > 0$ small enough, we obtain the Volterra inequality

$$u(t) \leq C_T \int_0^t |\zeta^\tau(t-s)|u(s)ds + \varepsilon u(t) + C_T \left(\int_0^T \mathbb{E} |\Delta M_s^V|^2 ds \right)^{1/2} + C_{T,\varepsilon}. \quad (\text{B.5})$$

Absorbing $\varepsilon u(t)$ into the left-hand side and applying the Volterra–Grönwall lemma (Lemma 3.2), using $\|\zeta^\tau\|_{L^1(0,T)} \leq C_T$, gives

$$u(t) \leq C_T \left(\int_0^T \mathbb{E} |\Delta M_s^V|^2 ds \right)^{1/2} + C_T, \quad t \in [0, T].$$

Since the additive constant C_T can be absorbed into the global constant (recall $u \geq 0$), we arrive at (B.4)

Step 2: Log-price stability. Write

$$\Delta X_T = -\frac{1}{2}c_X^2 \int_0^T \Delta V_s ds + c_X \int_0^T \left(\sqrt{\widehat{V}_s^\tau} - \sqrt{V_s^\tau} \right) dW_s^X + c_X \int_0^T \sqrt{\widehat{V}_s^\tau} d\Delta M_s^X.$$

Applying Cauchy-Schwarz inequality and BDG inequality and using (B.4) gives

$$\mathbb{E} \left[|\Delta X_T|^2 \right]^{1/2} \leq C_T \left(\int_0^T \mathbb{E} |\Delta M_s^X|^2 ds \right)^{1/2} + C_T \left(\int_0^T \mathbb{E} |\Delta M_s^V|^2 ds \right)^{1/2}. \quad (\text{B.6})$$

Define $\mathfrak{d}_\tau^2 := \int_0^T \mathbb{E} |\Delta M_s|^2 ds$. Then (B.6) implies

$$\mathbb{E} \left[|\Delta X_T|^2 \right]^{1/2} \leq C_T \mathfrak{d}_\tau. \quad (\text{B.7})$$

Step 3: From ΔX_T to prices. For each $(u, k) \in K$ inside the moment strip, the damped pricing functional $\text{Price}_{u,k}$ is locally Lipschitz in L^2 : there exists C_K such that

$$\left| \text{Price}_{u,k} \left(e^{\widehat{X}_T^\tau} \right) - \text{Price}_{u,k} \left(e^{X_T^\tau} \right) \right| \leq C_K \mathbb{E} \left[|\Delta X_T|^2 \right]^{1/2}. \quad (\text{B.8})$$

Combining (B.7)-(B.8) yields

$$\sup_{(u,k) \in K} \left| \text{Price}_{u,k} \left(\widehat{S}_T^\tau \right) - \text{Price}_{u,k} \left(S_T^\tau \right) \right| \leq C_{T,K} \mathfrak{d}_\tau.$$

By the microscopic invariance assumption, $\mathfrak{d}_\tau^2 = \int_0^T \mathbb{E} |M_t^\tau - W_t|^2 dt \leq C_T \tau^{-1/2}$; hence $\mathfrak{d}_\tau \leq C_T \tau^{-1/4}$ and the stated bound follows.

In particular. For $\alpha > \frac{1}{2}$ we have $\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)} \lesssim \tau^{\frac{1}{2}-\alpha}$. Thus the overall weak error is bounded by

$$\underbrace{C_{T,K} \tau^{-1/4}}_{\text{microscopic}} + \underbrace{C_{T,K} \tau^{\frac{1}{2}-\alpha}}_{\text{kernel}}.$$

If $\alpha \in (\frac{1}{2}, \frac{3}{4})$, then $\frac{1}{2} - \alpha \in (0, \frac{1}{4})$ and the kernel term dominates; the overall rate is $\mathcal{O}(\tau^{\frac{1}{2}-\alpha})$. If $\alpha \in [\frac{3}{4}, 1)$, then $\frac{1}{2} - \alpha \leq -\frac{1}{4}$ and the microscopic term dominates; the overall rate is $\mathcal{O}(\tau^{-1/4})$.

B.7 Proof of Proposition 3.11

Throughout the proof fix $T > 0$ and a complex number z with $\text{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$ as in Corollary 3.6. Write convolutions on $[0, T]$ as $(f * g)(t) := \int_0^t f(t-s)g(s)ds$, and abbreviate

$$\Delta \mathcal{A}(t) := \mathcal{A}^\tau(t; z) - \mathcal{A}(t; z), \quad \Delta \mathcal{B}(t) := \mathcal{B}^\tau(t; z) - \mathcal{B}(t; z).$$

By the affine–Volterra structure, the one-dimensional Riccati equations (defined in (3.1) and (3.3)) can be written as

$$\mathcal{A} = f_{\alpha,\gamma} * G(\mathcal{A}; z), \quad (\text{B.9})$$

$$\mathcal{B} = f_{\alpha,\gamma} * I(\mathcal{A}; z), \quad (\text{B.10})$$

and

$$\mathcal{A}^\tau = \zeta^\tau * G(\mathcal{A}^\tau; z), \quad (\text{B.11})$$

$$\mathcal{B}^\tau = \zeta^\tau * I(\mathcal{A}^\tau; z), \quad (\text{B.12})$$

where, for $h \in \mathbb{C}$,

$$G(h; z) := \frac{1}{2}(z^2 - z) + (\rho\xi z - \kappa)h + \frac{1}{2}\xi^2 h^2, \quad I(h; z) := \kappa\theta h. \quad (\text{B.13})$$

We introduce the auxiliary unknown H solving $H = f_{\alpha,\gamma} * G(H; z)$, whose existence and uniqueness is guaranteed by (Abi Jaber et al., 2019, Thm. B.1). Then one has

$$\mathcal{A} = f_{\alpha,\gamma} * H \quad \text{and} \quad \mathcal{B}(t; z) = \int_0^t I(H(s; z); z) ds.$$

The estimates below use only the boundedness and local Lipschitz properties of G and I ; hence either formulation yields the same bounds.

By Corollary 3.6 (well-posedness on the strip) there exist finite constants $K_{\mathcal{A}} = K_{\mathcal{A}}(T, z)$ and $K_I = K_I(T, z)$ such that

$$\sup_{t \in [0, T]} (|\mathcal{A}(t; z)| + |\mathcal{A}^\tau(t; z)|) \leq K_{\mathcal{A}}, \quad \sup_{t \in [0, T]} (|I(\mathcal{A}(t; z); z)| + |G(\mathcal{A}(t; z); z)|) \leq K_I, \quad (\text{B.14})$$

uniformly in τ (since $\sup_{\tau \geq 1} F^\tau(T) < \infty$ by Proposition 3.8). Moreover, by polynomial growth of the affine coefficients (defined in (B.13)) on compact z -sets, there are locally uniform Lipschitz constants $L_G = L_G(T, z)$ and $L_I = L_I(T, z)$ such that, for all $u, v \in \mathbb{C}$,

$$|G(u; z) - G(v; z)| \leq L_G(1 + |u| + |v|)|u - v|, \quad |I(u; z) - I(v; z)| \leq L_I|u - v|. \quad (\text{B.15})$$

Step 1: Stability for \mathcal{A} . Subtracting (B.9) and (B.11) gives

$$\Delta \mathcal{A}(t) = ((\zeta^\tau - f_{\alpha,\gamma}) * G(\mathcal{A}; z))(t) + (\zeta^\tau * (G(\mathcal{A}^\tau; z) - G(\mathcal{A}; z)))(t). \quad (\text{B.16})$$

Set $Y(t) := \sup_{0 \leq s \leq t} |\Delta \mathcal{A}(s)|$. Using (B.15)–(B.14) and Young’s convolution inequality in L^1 (and also in L^2 with Cauchy–Schwarz inequality), we obtain for all $t \in [0, T]$,

$$\begin{aligned} |\Delta \mathcal{A}(t)| &\leq \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0, T)} \sup_{s \leq t} |G(\mathcal{A}(s; z); z)| \\ &\quad + \int_0^t \zeta^\tau(t-s) L_G (1 + |\mathcal{A}^\tau(s; z)| + |\mathcal{A}(s; z)|) |\Delta \mathcal{A}(s)| ds \\ &\leq K_I \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0, T)} + L_G(1 + 2K_{\mathcal{A}}) (\zeta^\tau * Y)(t). \end{aligned} \quad (\text{B.17})$$

Equivalently,

$$Y(t) \leq a_\tau + b(\zeta^\tau * Y)(t), \quad a_\tau := K_I \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0, T)}, \quad b := L_G(1 + 2K_{\mathcal{A}}).$$

We now apply the Volterra–Grönwall inequality (see Lemma 3.2): if $k \geq 0$ with finite mass $K(t) := \int_0^t k(s) ds$, and $y(t) \leq a + (k * y)(t)$ on $[0, T]$, then

$$\sup_{t \leq T} y(t) \leq a \exp(K(T)). \quad (\text{B.18})$$

Here $k = \zeta^\tau$ so that $K(T) = F^\tau(T)$. Using (B.18) we deduce

$$\sup_{t \leq T} |\Delta \mathcal{A}(t)| \leq a_\tau \exp(bF^\tau(T)) \lesssim_{T,z} \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T)}. \quad (\text{B.19})$$

Repeating the first-term estimate in (B.17) with Cauchy–Schwarz inequality gives the alternative bound

$$\sup_{t \leq T} |\Delta \mathcal{A}(t)| \lesssim_{T,z} \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)},$$

and combining the L^1 bound (B.19) with the L^2 bound derived in the preceding line, we obtain

$$\sup_{t \leq T} |\Delta \mathcal{A}(t)| \leq C_T(z) \left(\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T)} + \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)} \right), \quad (\text{B.20})$$

where $C_T(z)$ is locally uniform on the strip $\{z \in \mathbb{C} : \operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]\}$ (since $F^\tau(T)$ is bounded in τ and K_A, K_I, L_G are locally uniform by Corollary 3.6).

Step 2: Stability for \mathcal{B} . Subtracting (B.10) and (B.12) yields

$$\Delta \mathcal{B}(t) = ((\zeta^\tau - f_{\alpha,\gamma}) * I(\mathcal{A}; z))(t) + (\zeta^\tau * (I(\mathcal{A}^\tau; z) - I(\mathcal{A}; z)))(t).$$

By (B.15)–(B.14) and Young’s inequality,

$$\begin{aligned} |\Delta \mathcal{B}(t)| &\leq \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T)} \sup_{s \leq t} |I(\mathcal{A}(s; z); z)| + L_I \int_0^t \zeta^\tau(t-s) |\Delta \mathcal{A}(s)| ds \\ &\leq K_I \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T)} + L_I F^\tau(T) \sup_{s \leq t} |\Delta \mathcal{A}(s)|. \end{aligned}$$

Taking the supremum over $t \in [0, T]$ and invoking (B.20) gives

$$\sup_{t \leq T} |\Delta \mathcal{B}(t)| \leq C_T(z) \left(\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T)} + \|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)} \right), \quad (\text{B.21})$$

with the same local uniformity on the strip $\{z \in \mathbb{C} : \operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]\}$.

Step 3: Incorporating the cumulative-kernel term. Since $\sup_{\tau \geq 1} F^\tau(T) < \infty$ by Proposition 3.8, the constants implicit in (B.20)–(B.21) are uniform in τ . Enlarging the right-hand sides by $\|F^\tau - F\|_{L^\infty(0,T)}$ keeps the inequalities valid, and by Proposition 3.8 this additional term is of the same order as $\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T) \cap L^2(0,T)}$.

Combining (B.20) and (B.21) yields the claimed estimate

$$\sup_{t \in [0, T]} (|\mathcal{A}^\tau(t; z) - \mathcal{A}(t; z)| + |\mathcal{B}^\tau(t; z) - \mathcal{B}(t; z)|) \leq C_T(z) \left(\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T) \cap L^2(0,T)} + \|F^\tau - F\|_{L^\infty(0,T)} \right).$$

Finally, from Proposition 3.8, we obtain

$$\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^1(0,T) \cap L^2(0,T)} + \|F^\tau - F\|_{L^\infty(0,T)} = \mathcal{O}\left(\tau^{\frac{1}{2}-\alpha}\right) \quad \text{as } \tau \rightarrow \infty.$$

B.8 Proof of Proposition 3.12

Fix $T > 0$ and $z \in \mathbb{C}$ with $\operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$ as in Corollary 3.6. Write

$$\Phi(T, z) := \mathcal{B}(T, z) + V_0 \mathcal{A}(T, z), \quad \Phi^\tau(T, z) := \mathcal{B}^\tau(T, z) + V_0 \mathcal{A}^\tau(T, z),$$

so that $\phi_T(z) = \exp(\Phi(T, z))$ and $\phi_T^\tau(z) = \exp(\Phi^\tau(T, z))$. By Corollary 3.6, both exponents are finite and

$$\sup_{\tau \geq 1} \max\{\operatorname{Re}(\Phi^\tau(T, z)), \operatorname{Re}(\Phi(T, z))\} =: M_T(z) < \infty.$$

Using the fundamental theorem of calculus along the line segment $\theta \mapsto \Phi(T, z) + \theta(\Phi^\tau(T, z) - \Phi(T, z))$, we obtain

$$\phi_T^\tau(z) - \phi_T(z) = \exp(\Phi(T, z)) \int_0^1 \exp(\theta(\Phi^\tau - \Phi)(T, z)) (\Phi^\tau - \Phi)(T, z) d\theta,$$

which implies that

$$\begin{aligned} |\phi_T^\tau(z) - \phi_T(z)| &\leq \exp\left(\sup_{\theta \in [0,1]} \operatorname{Re}\{\Phi(T, z) + \theta(\Phi^\tau - \Phi)(T, z)\}\right) |(\Phi^\tau - \Phi)(T, z)| \\ &\leq \exp(\max\{\operatorname{Re}(\Phi^\tau(T, z)), \operatorname{Re}(\Phi(T, z))\}) (|\mathcal{B}^\tau - \mathcal{B}|(T, z) + V_0 |\mathcal{A}^\tau - \mathcal{A}|(T, z)) \\ &\leq e^{M_T(z)} (|\mathcal{B}^\tau(T, z) - \mathcal{B}(T, z)| + V_0 |\mathcal{A}^\tau(T, z) - \mathcal{A}(T, z)|). \end{aligned}$$

Setting $C_T(z) := e^{M_T(z)}$ yields the claimed estimate. By Corollary 3.6, $M_T(z)$ is locally bounded on the strip $\{z \in \mathbb{C} : \operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]\}$, and hence so is $C_T(z)$.

B.9 Proof of Proposition 3.13

Since $\mathbb{E}[S_T] < \infty$, Jensen's inequality for the concave function $x \mapsto \sqrt{x}$ gives

$$\mathbb{E}[e^{\frac{1}{2}P_T}] = \mathbb{E}\left[\left(\frac{S_T}{K}\right)^{1/2}\right] = K^{-1/2} \mathbb{E}[S_T^{1/2}] \leq K^{-1/2} \sqrt{\mathbb{E}[S_T]} < \infty.$$

Hence, for every $u \in \mathbb{R}$,

$$\left|\phi_{P,T}\left(\frac{1}{2} + iu\right)\right| = \left|\mathbb{E}\left[e^{(\frac{1}{2} + iu)P_T}\right]\right| \leq \mathbb{E}\left[e^{\frac{1}{2}P_T}\right],$$

and therefore

$$\int_{\mathbb{R}} \frac{|\phi_{P,T}(\frac{1}{2} + iu)|}{\frac{1}{4} + u^2} du \leq \mathbb{E}\left[e^{\frac{1}{2}P_T}\right] \int_{\mathbb{R}} \frac{du}{\frac{1}{4} + u^2} = 2\pi \mathbb{E}\left[e^{\frac{1}{2}P_T}\right] < \infty.$$

This justifies exchanging expectation and integration below.

For all $y \in \mathbb{R}$,

$$\frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{(\frac{1}{2} + iu)y}}{\frac{1}{4} + u^2} du = \min\{1, e^y\}. \quad (\text{B.22})$$

Equivalently,

$$(e^y - 1)^+ = e^y - \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{(\frac{1}{2} + iu)y}}{\frac{1}{4} + u^2} du, \quad (\text{B.23})$$

which follows from the standard Fourier integral $\frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{iuy}}{a^2 + u^2} du = \frac{1}{2a} e^{-a|y|}$ with $a = \frac{1}{2}$.

Apply (B.23) with $y = P_T$ and exchange expectation and integral:

$$\mathbb{E}\left[(e^{P_T} - 1)^+\right] = \mathbb{E}\left[e^{P_T}\right] - \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\phi_{P,T}(\frac{1}{2} + iu)}{\frac{1}{4} + u^2} du.$$

Multiplying by K yields

$$C(T, K) = K \mathbb{E}[e^{P_T}] - \frac{K}{2\pi} \int_{\mathbb{R}} \frac{\phi_{P,T}(\frac{1}{2} + iu)}{\frac{1}{4} + u^2} du. \quad (\text{B.24})$$

Since $e^{P_T} = S_T/K$, we have $K \mathbb{E}[e^{P_T}] = \mathbb{E}[S_T]$; under the (undiscounted) risk-neutral measure this equals S_0 . The same representation holds for $C^{\text{aux}, \tau}(T, K)$ with $\phi_{P,T}^\tau$.

Subtract the two identities in (B.24); the $\mathbb{E}[S_T]$ (or S_0) terms cancel, and by the triangle inequality,

$$|C^{\text{aux}, \tau}(T, K) - C(T, K)| \leq \frac{K}{2\pi} \int_{\mathbb{R}} \frac{|\phi_{P,T}^\tau(\frac{1}{2} + iu) - \phi_{P,T}(\frac{1}{2} + iu)|}{\frac{1}{4} + u^2} du.$$

With $X_T := \log S_T$ and $\phi_{X,T}(z) := \mathbb{E}[e^{zX_T}]$, note that $\phi_{P,T}(z) = K^{-z}\phi_{X,T}(z)$; at $z = \frac{1}{2} + iu$ this gives $|\Delta\phi_{P,T}(z)| = K^{-1/2} |\Delta\phi_{X,T}(z)|$. Therefore

$$\frac{K}{2\pi} \int_{\mathbb{R}} \frac{|\Delta\phi_{P,T}(\frac{1}{2} + iu)|}{\frac{1}{4} + u^2} du = \frac{\sqrt{K}}{2\pi} \int_{\mathbb{R}} \frac{|\Delta\phi_{X,T}(\frac{1}{2} + iu)|}{\frac{1}{4} + u^2} du,$$

which yields the stated equivalent bound.

B.10 Proof of Theorem 3.14

Fix $T > 0$ and $K > 0$. On the line $\operatorname{Re}(z) = \frac{1}{2}$, Proposition 3.12 gives

$$|\phi_T^\tau(z) - \phi_T(z)| \leq C_T(z) (|\mathcal{B}^\tau(T, z) - \mathcal{B}(T, z)| + V_0 |\mathcal{A}^\tau(T, z) - \mathcal{A}(T, z)|).$$

By Proposition 3.11 and Proposition 3.8,

$$\sup_{t \leq T} (|\mathcal{B}^\tau(t; z) - \mathcal{B}(t; z)| + |\mathcal{A}^\tau(t; z) - \mathcal{A}(t; z)|) \leq C_T \tau^{\frac{1}{2} - \alpha},$$

with constants uniform on the line $\operatorname{Re}(z) = \frac{1}{2}$ thanks to Corollary 3.6 (the latter provides a uniform bound on $\operatorname{Re}(\mathcal{B} + V_0 \mathcal{A})$ over the strip $\{z \in \mathbb{C} : \operatorname{Re}(z) \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]\}$, and hence on $C_T(z)$). Therefore

$$\left| \phi_T^\tau\left(\frac{1}{2} + iu\right) - \phi_T\left(\frac{1}{2} + iu\right) \right| \leq C_T \tau^{\frac{1}{2} - \alpha} \quad \text{for all } u \in \mathbb{R}.$$

Applying the Carr–Madan map in the moneyness variable (Proposition 3.13) yields

$$|C^{\text{aux}, \tau}(T, K) - C(T, K)| \leq \frac{K}{2\pi} \int_{\mathbb{R}} \frac{|\phi_T^\tau(\frac{1}{2} + iu) - \phi_T(\frac{1}{2} + iu)|}{\frac{1}{4} + u^2} du \leq C_{T,K} \tau^{\frac{1}{2} - \alpha},$$

since $\int_{\mathbb{R}} (\frac{1}{4} + u^2)^{-1} du < \infty$.

Let $C^{\text{INAR}, \tau}(T, K)$ be the price in the microscopic model. By the triangle inequality and Theorem 3.10,

$$\begin{aligned} |C^{\text{INAR}, \tau}(T, K) - C(T, K)| &\leq |C^{\text{INAR}, \tau}(T, K) - C^{\text{aux}, \tau}(T, K)| + |C^{\text{aux}, \tau}(T, K) - C(T, K)| \\ &\leq C_{T,K} \left(\tau^{-1/4} + \tau^{\frac{1}{2} - \alpha} \right), \end{aligned}$$

and the stated piecewise rates follow by comparing the two powers.

B.11 Proof of Lemma 3.15

Since the map $u \mapsto u^+$ is 1-Lipschitz, we have

$$|(a - K)^+ - (b - K)^+| \leq |a - b|, \quad \text{for any } a, b \in \mathbb{R}.$$

Applying this with $a = \frac{1}{T} \int_0^T S_t^\tau dt$ and $b = \frac{1}{T} \int_0^T S_t dt$, we obtain

$$\begin{aligned} |AA^{\text{aux}, \tau}(T, K) - AA(T, K)| &= \left| \mathbb{E} \left[\left(\frac{1}{T} \int_0^T S_t^\tau dt - K \right)^+ - \left(\frac{1}{T} \int_0^T S_t dt - K \right)^+ \right] \right| \\ &\leq \mathbb{E} \left[\left| \frac{1}{T} \int_0^T (S_t^\tau - S_t) dt \right| \right] \\ &\leq \frac{1}{T} \int_0^T \mathbb{E} [|S_t^\tau - S_t|] dt, \end{aligned}$$

where the last inequality follows from Jensen's inequality (for the convex function $x \mapsto |x|$) together with Tonelli's theorem. This proves the claim.

B.12 Proof of Lemma 3.16

Step 1: Exponential moments of $\int_0^t V_s ds$. This step was rigorously proven in the proof of Lemma 3.3 (Appendix B.2). We established that for

$$\Theta_T := \frac{2}{c_V^2 \|\Phi\|_{L^\infty(0,T)}^2} \quad \text{and} \quad \Theta_{\text{eff}} := \Theta_T/4,$$

the moment generating function $M_t(\lambda) := \mathbb{E} \exp\left(\lambda \int_0^t V_s ds\right)$ satisfies the recursion $(M_T(\lambda))^2 \leq e^{2\lambda D_T} M_T(4c\lambda^2)$, where $c = 1/(2\Theta_T)$. This implies that for any $\lambda \in [0, \Theta_{\text{eff}})$,

$$\sup_{t \in [0, T]} \mathbb{E} \exp\left(\lambda \int_0^t V_s ds\right) \leq M_T(\lambda) < \infty. \quad (\text{B.25})$$

Step 2: Moments of e^{yX_t} . This step follows by applying the similar argument as in Step 2 of the proof of Lemma 3.3 (Appendix B.2). The log-price satisfies $dX_t = -\frac{1}{2}c_X^2 V_t dt + c_X \sqrt{V_t} dW_t^X$. For $y \geq 0$, we showed using the Cauchy-Schwarz inequality that

$$\mathbb{E}[e^{yX_t}] \leq C_y \left(\mathbb{E} \exp\left((2y^2 - y) c_X^2 \int_0^t V_s ds\right) \right)^{1/2},$$

where $C_y := e^{yX_0} < \infty$. Let $\lambda_y := c_X^2(2y^2 - y)$. The expectation $\mathbb{E}[e^{yX_t}]$ is finite, provided $\lambda_y < \Theta_{\text{eff}}$. This is precisely the condition (3.14) in the lemma statement. If this condition holds, then by (B.25),

$$\sup_{t \in [0, T]} \mathbb{E}[e^{yX_t}] \leq C_y \sup_{t \in [0, T]} \sqrt{M_t(\lambda_y)} \leq C_y \sqrt{M_T(\lambda_y)} < \infty,$$

which proves part (a) of the lemma. The formulas for y_{\max} and θ^* follow from solving the quadratic inequality $c_X^2(2y^2 - y) < \Theta_{\text{eff}}$ for $y \geq 0$.

Step 3: Uniform-in- τ prelimit. For part (b), we use the proof of Proposition 3.5 (Appendix B.3). We defined the uniform thresholds:

$$\Theta^* := \frac{2}{c_V^2 \sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0,T)}^2} \quad \text{and} \quad \Theta_{\text{eff}}^* := \Theta^*/4.$$

The exact same logic from Step 2 applies, replacing Θ_{eff} with Θ_{eff}^* . For every $y \geq 0$ satisfying $c_X^2(2y^2 - y) < \Theta_{\text{eff}}^*$, we have

$$\sup_{\tau \geq 1} \sup_{t \in [0, T]} \mathbb{E} \left[e^{yX_t^{(\tau)}} \right] \leq C_y \sup_{\tau \geq 1} \sqrt{M_\tau(\lambda_y)} < \infty,$$

which proves part (b).

Step 4: Coarse sufficient conditions. For part (c), we use the Cauchy-Schwarz inequality $\|\Phi\|_{L^\infty(0,T)} \leq \sqrt{T} \|\Phi\|_{L^2(0,T)}$, which implies

$$\|\Phi\|_{L^\infty(0,T)}^2 \leq T \int_0^T (f_{\alpha, \gamma}(s))^2 ds = T \mathcal{G}(T).$$

We substitute this into the definition of Θ_T :

$$\Theta_T = \frac{2}{c_V^2 \|\Phi\|_{L^\infty(0,T)}^2} \geq \frac{2}{c_V^2 T \mathcal{G}(T)}.$$

The effective threshold from Step 1 is $\Theta_{\text{eff}} = \Theta_T/4$. Therefore,

$$\Theta_{\text{eff}} \geq \frac{1}{2c_V^2 T \mathcal{G}(T)}.$$

The condition for the limit model in part (a) is $c_X^2(2y^2 - y) < \Theta_{\text{eff}}$. A sufficient condition for this is thus:

$$c_X^2(2y^2 - y) < \frac{1}{2c_V^2 T G(T)},$$

which rearranges to the first statement in part (c):

$$c_X^2 c_V^2 T \left(\int_0^T (f_{\alpha, \gamma}(s))^2 ds \right) (2y^2 - y) < \frac{1}{2}.$$

The same logic applies to the uniform-in- τ case. We have

$$\sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0, T)}^2 \leq T \left(\sup_{\tau \geq 1} \int_0^T (\zeta^\tau(s))^2 ds \right) = T \sup_{\tau \geq 1} G^\tau(T).$$

The uniform threshold from Step 3 is $\Theta^* = \frac{2}{c_V^2 \sup_{\tau \geq 1} \|\Phi^\tau\|_{L^\infty(0, T)}^2}$, which implies

$$\Theta^* \geq \frac{2}{c_V^2 T \sup_{\tau \geq 1} G^\tau(T)}.$$

The effective uniform threshold is $\Theta_{\text{eff}}^* = \Theta^*/4$, so

$$\Theta_{\text{eff}}^* \geq \frac{1}{2c_V^2 T \sup_{\tau \geq 1} G^\tau(T)}.$$

The condition for the uniform prelimit model in part (b) is $c_X^2(2y^2 - y) < \Theta_{\text{eff}}^*$. A sufficient condition for this is:

$$c_X^2(2y^2 - y) < \frac{1}{2c_V^2 T \sup_{\tau \geq 1} G^\tau(T)},$$

which rearranges to the second statement in part (c):

$$c_X^2 c_V^2 T \left(\sup_{\tau \geq 1} \int_0^T (\zeta^\tau(s))^2 ds \right) (2y^2 - y) < \frac{1}{2}.$$

The final claim regarding the canonical case $c_X = 1, c_V = \nu$ follows directly. This completes the proof of part (c).

B.13 Proof of Theorem 3.18

By Lemma 3.16 with $y = 2$ we have the order-2 exponential moments

$$\sup_{t \in [0, T]} \mathbb{E} [e^{2X_t}] \leq C_{\text{exp}}, \quad \sup_{\tau \geq 1} \sup_{t \in [0, T]} \mathbb{E} [e^{2X_t^\tau}] \leq C_{\text{exp}}, \quad (\text{B.26})$$

provided the (equivalent) parameter thresholds in Lemma 3.16 hold (e.g. $c_X^2 < \Theta_T$ and $c_X^2 < \Theta^*$ for the prelimit, or the coarse sufficient inequalities there). We now prove the two claims.

By Lemma 3.15,

$$|AA^{\text{aux}, \tau}(T, K) - AA(T, K)| \leq \frac{1}{T} \int_0^T \mathbb{E} |S_t^\tau - S_t| dt, \quad S_t^\tau = e^{X_t^\tau}, \quad S_t = e^{X_t}. \quad (\text{B.27})$$

Using $|e^x - e^y| \leq (e^x + e^y)|x - y|$ and Cauchy-Schwarz inequality,

$$\mathbb{E} |S_t^\tau - S_t| \leq \left(\mathbb{E} \left[\left(e^{X_t^\tau} + e^{X_t} \right)^2 \right] \right)^{1/2} (\mathbb{E} |X_t^\tau - X_t|^2)^{1/2}.$$

By (B.26) (which provides bounds uniform in t and τ), we have

$$\sup_{t \in [0, T], \tau \geq 1} \mathbb{E} \left[\left(e^{X_t^\tau} + e^{X_t} \right)^2 \right] \leq 2 \sup_{t \in [0, T], \tau \geq 1} \mathbb{E}[e^{2X_t^\tau}] + 2 \sup_{t \in [0, T]} \mathbb{E}[e^{2X_t}] \leq 4C_{\text{exp}}.$$

Hence, this uniform bound holds for any t and τ , allowing us to write

$$\mathbb{E}|S_t^\tau - S_t| \leq 2\sqrt{C_{\text{exp}}} \left(\mathbb{E}|X_t^\tau - X_t|^2 \right)^{1/2}, \quad t \in [0, T]. \quad (\text{B.28})$$

Integrate (B.28) over t and apply Cauchy–Schwarz inequality in time:

$$\frac{1}{T} \int_0^T \mathbb{E}|S_t^\tau - S_t| dt \leq \frac{2\sqrt{C_{\text{exp}}}}{\sqrt{T}} \left(\int_0^T \mathbb{E}|X_t^\tau - X_t|^2 dt \right)^{1/2}.$$

By the standard Volterra L^2 stability (BDG inequality and Young convolution) and Proposition 3.8,

$$\int_0^T \mathbb{E}|X_t^\tau - X_t|^2 dt \leq C_T \|\zeta^\tau - f_{\alpha, \gamma}\|_{L^2(0, T)}^2 \lesssim_T \tau^{1-2\alpha}.$$

Combining with (B.27) yields

$$|AA^{\text{aux}, \tau}(T, K) - AA(T, K)| \leq C_{T, K} \tau^{\frac{1}{2} - \alpha}.$$

Next, we estimate the microscopic error $|AA^{\text{INAR}, \tau}(T, K) - AA^{\text{aux}, \tau}(T, K)|$. Let \bar{S}_{INAR} and \bar{S}_{aux} be the time-averaged prices for the microscopic and auxiliary SDE models, respectively. Then, we have

$$\begin{aligned} |AA^{\text{INAR}, \tau}(T, K) - AA^{\text{aux}, \tau}(T, K)| &\leq \mathbb{E} [|\bar{S}_{\text{INAR}} - \bar{S}_{\text{aux}}|] \\ &\leq \frac{1}{T} \int_0^T \mathbb{E} [|\hat{S}_t^\tau - S_t^\tau|] dt, \end{aligned}$$

where the first step uses the 1-Lipschitz property of the payoff and the second uses Fubini's theorem and Jensen's inequality. We bound the integrand $\mathbb{E} [|\hat{S}_t^\tau - S_t^\tau|]$ using the inequality $|e^x - e^y| \leq (e^x + e^y)|x - y|$ and the Cauchy-Schwarz inequality:

$$\mathbb{E} [|\hat{S}_t^\tau - S_t^\tau|] \leq \left(\mathbb{E} \left[\left(e^{\hat{X}_t^\tau} + e^{X_t^\tau} \right)^2 \right] \right)^{1/2} \left(\mathbb{E} [|\hat{X}_t^\tau - X_t^\tau|^2] \right)^{1/2}.$$

By the uniform exponential moment bound (B.26), $\mathbb{E} \left[\left(e^{\hat{X}_t^\tau} + e^{X_t^\tau} \right)^2 \right] \leq 4C_{\text{exp}}$. Thus,

$$\mathbb{E} [|\hat{S}_t^\tau - S_t^\tau|] \leq 2\sqrt{C_{\text{exp}}} \left(\mathbb{E} [|\hat{X}_t^\tau - X_t^\tau|^2] \right)^{1/2}.$$

Substituting this back into the integral for the Asian price error and applying the Cauchy-Schwarz inequality in the time variable t :

$$\begin{aligned} |AA^{\text{INAR}, \tau}(T, K) - AA^{\text{aux}, \tau}(T, K)| &\leq \frac{2\sqrt{C_{\text{exp}}}}{T} \int_0^T \left(\mathbb{E} [|\hat{X}_t^\tau - X_t^\tau|^2] \right)^{1/2} dt \\ &\leq \frac{2\sqrt{C_{\text{exp}}}}{T} \left(\int_0^T 1^2 dt \right)^{1/2} \left(\int_0^T \mathbb{E} [|\hat{X}_t^\tau - X_t^\tau|^2] dt \right)^{1/2} \\ &= \frac{2\sqrt{C_{\text{exp}}}}{\sqrt{T}} \left(\int_0^T \mathbb{E} [|\hat{X}_t^\tau - X_t^\tau|^2] dt \right)^{1/2}. \end{aligned}$$

The time-integrated L^2 log-price error is controlled by the L^2 stability of the Volterra SDE, which in turn is driven by the time-integrated L^2 error of the drivers (see proof of Theorem 3.10):

$$\int_0^T \mathbb{E} \left[\left| \hat{X}_t^\tau - X_t^\tau \right|^2 \right] dt \leq C_T \int_0^T \mathbb{E} |M_t^\tau - W_t|^2 dt.$$

By Proposition 3.9, the driver error is bounded by $\int_0^T \mathbb{E} |M_t^\tau - W_t|^2 dt \leq C_T \tau^{-1/2}$. Combining these results yields:

$$|AA^{\text{INAR},\tau}(T, K) - AA^{\text{aux},\tau}(T, K)| \leq \frac{C_T}{\sqrt{T}} \left(C_T \tau^{-1/2} \right)^{1/2} = C_{T,K} \tau^{-1/4}.$$

Finally, combining both error terms via the triangle inequality:

$$\begin{aligned} |AA^{\text{INAR},\tau}(T, K) - AA(T, K)| &\leq |AA^{\text{INAR},\tau}(T, K) - AA^{\text{aux},\tau}(T, K)| + |AA^{\text{aux},\tau}(T, K) - AA(T, K)| \\ &\leq C_{T,K} \left(\tau^{-1/4} + \tau^{\frac{1}{2}-\alpha} \right), \end{aligned}$$

where the constant $C_{T,K}$ is locally uniform in K and in model parameters on compact sets satisfying the stated exponential-moment condition. This completes the proof.

B.14 Proof of Proposition 3.20

Write $M := \sup_{0 \leq t \leq T} X_t$ and $M^\tau := \sup_{0 \leq t \leq T} X_t^\tau$, so that $LB(T, K) = \mathbb{E} \left[(e^M - K)^+ \right]$ and $LB^{\text{aux},\tau}(T, K) = \mathbb{E} \left[(e^{M^\tau} - K)^+ \right]$. Fix a mesh $h \in (0, T]$, let $N := \lceil T/h \rceil$ and $t_k := \min\{kh, T\}$, and denote

$$M_h := \max_{1 \leq k \leq N} X_{t_k}, \quad M_h^\tau := \max_{1 \leq k \leq N} X_{t_k}^\tau.$$

By continuity of X and X^τ and the moment strip (which yields standard quadratic-moment bounds on increments via BDG inequality), there exists $C_T < \infty$ such that

$$\mathbb{E} [|M - M_h|] \leq C_T h^{1/2}, \quad \mathbb{E} [|M^\tau - M_h^\tau|] \leq C_T h^{1/2}. \quad (\text{B.29})$$

(Indeed, by BDG inequality, the discretization error of the supremum is bounded as $\mathbb{E}[\sup_{t \in [0, T]} |X_t - X_{t^{(h)}}|] \lesssim h^{1/2}$, where $t^{(h)}$ is the nearest grid time to t ; similarly for X^τ .)

Since $u \mapsto (e^u - K)^+$ is 1-Lipschitz in u up to the multiplicative factor e^θ for some θ between the arguments, we use the elementary bound

$$|(e^a - K)^+ - (e^b - K)^+| \leq e^{\max\{a, b\}} |a - b| \quad (a, b \in \mathbb{R}),$$

and Lemma 3.16 (which yields exponential moments beyond order 1 uniformly in $t \leq T$ and in τ) to deduce

$$\sup_{t \leq T} \mathbb{E} [e^{X_t}] < \infty \quad \text{and} \quad \sup_{\tau \geq 1} \sup_{t \leq T} \mathbb{E} [e^{X_t^\tau}] < \infty.$$

Then, by Cauchy-Schwarz inequality,

$$\left| \mathbb{E} \left[(e^M - K)^+ \right] - \mathbb{E} \left[(e^{M_h} - K)^+ \right] \right| + \left| \mathbb{E} \left[(e^{M^\tau} - K)^+ \right] - \mathbb{E} \left[(e^{M_h^\tau} - K)^+ \right] \right| \leq C_{T,K} h^{1/2}. \quad (\text{B.30})$$

Since the running maximum is a non-smooth (non-Lipschitz) path functional, we replace it by a log-sum-exp “soft-max” with explicit bias and a 1-Lipschitz property compatible with our L^2 stability bounds. A convenient smoothing of the running maximum of the log-price is the continuous-time log-sum-exp (“soft-max”) functional

$$\text{SM}_\varepsilon(X.) := \varepsilon \log \left(\int_0^T e^{X_t/\varepsilon} dt \right) - \varepsilon \log T, \quad \varepsilon \in (0, 1].$$

For $\varepsilon > 0$, define the *discrete, normalized soft-max*

$$\mathbf{SM}_\varepsilon^{(h)}(x.) := \varepsilon \log \left(\frac{1}{N} \sum_{k=1}^N e^{x_{t_k}/\varepsilon} \right). \quad (\text{B.31})$$

It is standard that: (i) $\max_k x_{t_k} - \varepsilon \log N \leq \mathbf{SM}_\varepsilon^{(h)}(x.) \leq \max_k x_{t_k}$; (ii) the map $x \mapsto \mathbf{SM}_\varepsilon^{(h)}(x.)$ is 1-Lipschitz w.r.t. the sup-norm on the grid:¹

$$\left| \mathbf{SM}_\varepsilon^{(h)}(x.) - \mathbf{SM}_\varepsilon^{(h)}(y.) \right| \leq \max_{1 \leq k \leq N} |x_{t_k} - y_{t_k}|. \quad (\text{B.32})$$

Consequently,

$$\left| \mathbb{E} \left[\left(e^{M_h^\tau} - K \right)^+ \right] - \mathbb{E} \left[\left(e^{M_h} - K \right)^+ \right] \right| \leq A_1(h, \varepsilon) + A_2(h, \varepsilon),$$

where

$$\begin{aligned} A_1 := & \left| \mathbb{E} \left[\left(e^{M_h^\tau} - K \right)^+ \right] - \mathbb{E} \left[\left(e^{\mathbf{SM}_\varepsilon^{(h)}(X^\tau) + \varepsilon \log N} - K \right)^+ \right] \right| \\ & + \left| \mathbb{E} \left[\left(e^{M_h} - K \right)^+ \right] - \mathbb{E} \left[\left(e^{\mathbf{SM}_\varepsilon^{(h)}(X) + \varepsilon \log N} - K \right)^+ \right] \right|, \end{aligned}$$

and

$$A_2 := \left| \mathbb{E} \left[\left(e^{\mathbf{SM}_\varepsilon^{(h)}(X^\tau) + \varepsilon \log N} - K \right)^+ \right] - \mathbb{E} \left[\left(e^{\mathbf{SM}_\varepsilon^{(h)}(X) + \varepsilon \log N} - K \right)^+ \right] \right|.$$

Since $\mathbf{SM}_\varepsilon^{(h)} \leq M_h \leq \mathbf{SM}_\varepsilon^{(h)} + \varepsilon \log N$,

$$\begin{aligned} 0 \leq (e^{M_h} - K)^+ - \left(e^{\mathbf{SM}_\varepsilon^{(h)} + \varepsilon \log N} - K \right)^+ & \leq e^{M_h} - e^{\mathbf{SM}_\varepsilon^{(h)} + \varepsilon \log N} \\ & \leq e^{M_h} \left(1 - e^{-(M_h - \mathbf{SM}_\varepsilon^{(h)} - \varepsilon \log N)} \right) \leq e^{M_h} \varepsilon \log N, \end{aligned}$$

using $1 - e^{-x} \leq x$ for $x \geq 0$. Taking expectations and applying the same bound to the second term in the definition of A_1 (the term involving X^τ) gives

$$A_1(h, \varepsilon) \leq \varepsilon \log N \left(\mathbb{E} [e^{M_h}] + \mathbb{E} [e^{M_h^\tau}] \right). \quad (\text{B.33})$$

We now claim that $\sup_{h \in (0, T]} (\mathbb{E} [e^{M_h}] + \sup_{\tau \geq 1} \mathbb{E} [e^{M_h^\tau}]) \leq C_T < \infty$. Indeed, write $\tilde{S}_t := e^{X_t}$. Then $(\tilde{S}_t)_{t \in [0, T]}$ is a positive martingale under the Novikov-type threshold (Lemma 3.3); moreover, by Lemma 3.16 there exists $p > 1$ with $\sup_{t \leq T} \mathbb{E} [\tilde{S}_t^p] < \infty$, and the same holds uniformly in τ for the prelimit models. By Doob's L^p inequality,

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} \tilde{S}_t \right] \leq \left(\mathbb{E} \left[\sup_{0 \leq t \leq T} \tilde{S}_t^p \right] \right)^{1/p} \leq \left(\frac{p}{p-1} \right) \left(\mathbb{E} [\tilde{S}_T^p] \right)^{1/p} \leq C_T.$$

Since $e^{M_h} = \sup_{1 \leq k \leq N} e^{X_{t_k}} \leq \sup_{0 \leq t \leq T} \tilde{S}_t$, we obtain $\sup_h \mathbb{E} [e^{M_h}] \leq C_T$, and similarly $\sup_{h, \tau} \mathbb{E} [e^{M_h^\tau}] \leq C_T$. Plugging this into (B.33) yields

$$A_1(h, \varepsilon) \leq C_T \varepsilon \log \frac{T}{h}. \quad (\text{B.34})$$

¹It is straightforward to check (i). Let us prove (ii). Let $f(x) = \mathbf{SM}_\varepsilon^{(h)}(x.)$ and let $\delta = \max_k |x_{t_k} - y_{t_k}|$. By definition, $x_{t_k} \leq y_{t_k} + \delta$ for all k . Since $u \mapsto e^{u/\varepsilon}$ is increasing, $e^{x_{t_k}/\varepsilon} \leq e^{(y_{t_k} + \delta)/\varepsilon} = e^{\delta/\varepsilon} e^{y_{t_k}/\varepsilon}$. Summing over k and dividing by N gives $\frac{1}{N} \sum_k e^{x_{t_k}/\varepsilon} \leq e^{\delta/\varepsilon} \left(\frac{1}{N} \sum_k e^{y_{t_k}/\varepsilon} \right)$. Applying the increasing function $\varepsilon \log(\cdot)$ to both sides yields $f(x) \leq \varepsilon \log(e^{\delta/\varepsilon}) + f(y) = \delta + f(y)$, so $f(x) - f(y) \leq \delta$. By symmetry, $f(y) - f(x) \leq \delta$. Combining these gives $|f(x) - f(y)| \leq \delta$.

By the mean-value bound $|e^u - e^v| \leq (e^u + e^v)|u - v|$ and (B.32),

$$A_2 \leq \mathbb{E} \left[\left(e^{\text{SM}_\varepsilon^{(h)}(X^\tau) + \varepsilon \log N} + e^{\text{SM}_\varepsilon^{(h)}(X) + \varepsilon \log N} \right) \left| \text{SM}_\varepsilon^{(h)}(X^\tau) - \text{SM}_\varepsilon^{(h)}(X) \right| \right] \leq C_T \mathbb{E} \left[\max_{k \leq N} |X_{t_k}^\tau - X_{t_k}| \right].$$

By Cauchy–Schwarz inequality and the exponential moment bounds from Lemma 3.16,

$$A_2 \leq C_T \left(\sum_{k=1}^N \mathbb{E} |X_{t_k}^\tau - X_{t_k}|^2 \right)^{1/2}.$$

We now bound the L^2 discretization error of the log-price, which is required to control A_2 . Set $\Delta X_t := X_t^\tau - X_t$. By subtracting the Volterra equations for X^τ and X one can see that $(\Delta X_t)_{t \in [0, T]}$ solves a linear Volterra-type system driven by the same Brownian motion. Applying BDG inequality to the stochastic integrals and using the Lipschitz properties of the coefficients, one obtains for $t \in [0, T]$ an inequality of the form

$$\mathbb{E} [|\Delta X_t|^2] \leq C_T \left(\int_0^t (\zeta^\tau(t-s) - f_{\alpha, \gamma}(t-s))^2 ds + \int_0^t (f_{\alpha, \gamma}(t-s))^2 \sup_{r \leq s} \mathbb{E} [|\Delta X_r|^2] ds \right). \quad (\text{B.35})$$

We claim

$$\sup_{t \in [0, T]} \mathbb{E} [|\Delta X_t|^2] \leq C_T \|\zeta^\tau - f_{\alpha, \gamma}\|_{L^2(0, T)}^2. \quad (\text{B.36})$$

In fact, define

$$u(t) := \sup_{r \leq t} \mathbb{E} [|\Delta X_r|^2], \quad t \in [0, T].$$

Then (B.35) implies

$$u(t) \leq C_T \int_0^t (\zeta^\tau(t-s) - f_{\alpha, \gamma}(t-s))^2 ds + C_T \int_0^t f_{\alpha, \gamma}(t-s)^2 u(s) ds.$$

For the first term we simply use that, for every $t \in [0, T]$,

$$\int_0^t (\zeta^\tau(t-s) - f_{\alpha, \gamma}(t-s))^2 ds = \int_0^t (\zeta^\tau(s) - f_{\alpha, \gamma}(s))^2 ds \leq \|\zeta^\tau - f_{\alpha, \gamma}\|_{L^2(0, T)}^2.$$

Hence

$$u(t) \leq a + \int_0^t b(t-s) u(s) ds, \quad t \in [0, T],$$

with

$$a := C_T \|\zeta^\tau - f_{\alpha, \gamma}\|_{L^2(0, T)}^2, \quad b(s) := C_T f_{\alpha, \gamma}(s)^2, \quad s \in [0, T].$$

Note that $b \in L^1(0, T)$ because $f_{\alpha, \gamma} \in L^2(0, T)$. We are thus exactly in the setting of Lemma 3.2, which yields

$$u(t) \leq a \exp \left(\int_0^t b(s) ds \right) \leq a \exp \left(C_T \int_0^T f_{\alpha, \gamma}(s)^2 ds \right), \quad t \in [0, T].$$

Since T and the model parameters are fixed, the exponential factor can be absorbed into a new constant $C_T < \infty$. This gives

$$\sup_{t \in [0, T]} u(t) = \sup_{t \in [0, T]} \sup_{r \leq t} \mathbb{E} [|\Delta X_r|^2] = \sup_{t \in [0, T]} \mathbb{E} [|\Delta X_t|^2] \leq C_T \|\zeta^\tau - f_{\alpha, \gamma}\|_{L^2(0, T)}^2,$$

which is (B.36).

In particular,

$$\frac{1}{N} \sum_{k=1}^N \mathbb{E} |X_{t_k}^\tau - X_{t_k}|^2 \leq \sup_{t \in [0, T]} \mathbb{E} [|\Delta X_t|^2] \leq C_T \|\zeta^\tau - f_{\alpha, \gamma}\|_{L^2(0, T)}^2.$$

By Proposition 3.8 we have the kernel-approximation rate $\|\zeta^\tau - f_{\alpha,\gamma}\|_{L^2(0,T)} \leq C_1 \tau^{\frac{1}{2}-\alpha}$, so that

$$\frac{1}{N} \sum_{k=1}^N \mathbb{E} |X_{t_k}^\tau - X_{t_k}|^2 \leq C'_T \tau^{1-2\alpha}. \quad (\text{B.37})$$

Recalling the definition of A_2 and combining (B.37) with the preceding estimate, we obtain

$$A_2 \leq C_T \left(\sum_{k=1}^N \mathbb{E} |X_{t_k}^\tau - X_{t_k}|^2 \right)^{1/2} = C_T (N \tau^{1-2\alpha})^{1/2} = C_T N^{1/2} \tau^{\frac{1}{2}-\alpha}.$$

Since $N = \lceil T/h \rceil \asymp h^{-1}$ for fixed T , this is equivalently

$$A_2 \lesssim_T h^{-1/2} \tau^{\frac{1}{2}-\alpha}. \quad (\text{B.38})$$

By combining (B.30)–(B.38), we obtain

$$|LB^{\text{aux},\tau}(T, K) - LB(T, K)| \leq C_{T,K} \left(h^{1/2} + \varepsilon \log \frac{T}{h} + h^{-1/2} \tau^{\frac{1}{2}-\alpha} \right).$$

Choose $\varepsilon := h$ and then balance the last two terms by $h^{1/2} = h^{-1/2} \tau^{\frac{1}{2}-\alpha}$, i.e. $h = \tau^{\frac{1}{2}-\alpha}$. Since $T \leq 1$ is fixed, $h \log \frac{T}{h} \lesssim h^{1/2}$ as $h \downarrow 0$; thus

$$|LB^{\text{aux},\tau}(T, K) - LB(T, K)| \leq C_{T,K} \left(h^{1/2} + h^{-1/2} \tau^{\frac{1}{2}-\alpha} \right) = C_{T,K} \tau^{\frac{1}{4}-\frac{\alpha}{2}}.$$

This proves the claim.

B.15 Proof of Theorem 3.21

By the triangle inequality,

$$|LB^{\text{INAR},\tau}(T, K) - LB(T, K)| \leq \underbrace{|LB^{\text{INAR},\tau}(T, K) - LB^{\text{aux},\tau}(T, K)|}_{(i)} + \underbrace{|LB^{\text{aux},\tau}(T, K) - LB(T, K)|}_{(ii)}.$$

Term (ii) is controlled by Proposition 3.20:

$$(ii) \leq C_{T,K} \tau^{\frac{1}{4}-\frac{\alpha}{2}}.$$

It remains to prove (i) $\lesssim \tau^{-1/8}$. Set $M := \sup_{t \leq T} X_t^\tau$, $\widehat{M} := \sup_{t \leq T} \widehat{X}_t^\tau$ for the auxiliary and microscopic models (same coefficients and kernel ζ^τ , drivers W and M^τ , respectively), so that $LB^{\text{aux},\tau}(T, K) = \mathbb{E} \left[(e^M - K)^+ \right]$ and $LB^{\text{INAR},\tau}(T, K) = \mathbb{E} \left[(e^{\widehat{M}} - K)^+ \right]$.

Fix a mesh $h \in (0, T]$, let $N := \lceil T/h \rceil$ and $t_k := \min\{kh, T\}$, and write

$$M_h := \max_{1 \leq k \leq N} X_{t_k}^\tau, \quad \widehat{M}_h := \max_{1 \leq k \leq N} \widehat{X}_{t_k}^\tau.$$

The bound on the discretization error is derived from the path regularity of the underlying processes. The error in approximating the continuous supremum M by the discrete maximum M_h is controlled by the process's modulus of continuity. For a stochastic process driven by Brownian motion, a standard result combining the BDG inequality with Kolmogorov's continuity criterion yields an estimate for the expected maximum fluctuation:

$$\mathbb{E} [|M - M_h|] \leq \mathbb{E} \left[\sup_{|t-s| \leq h, t, s \in [0, T]} |X_t - X_s| \right] \leq C_T \sqrt{h \log(T/h)}.$$

This bound relies on the fact that the volatility process in the model has finite moments (as established in Lemma 3.16), ensuring that the increments of X_t have sufficiently high moments. The same reasoning applies to the auxiliary process \widehat{X}_t . Since the payoff function $f(u) = (e^u - K)^+$ is locally Lipschitz, this pathwise error translates into an error in the expected payoff, which, after applying Cauchy-Schwarz inequality and using the exponential moment bounds from Lemma 3.16, establishes the overall discretization error rate:

$$\left| \mathbb{E} \left[(e^M - K)^+ \right] - \mathbb{E} \left[(e^{M_h} - K)^+ \right] \right| + \left| \mathbb{E} \left[(e^{\widehat{M}} - K)^+ \right] - \mathbb{E} \left[(e^{\widehat{M}_h} - K)^+ \right] \right| \leq C_{T,K} \sqrt{h \log(T/h)}. \quad (\text{B.39})$$

Recall the discrete soft-max (B.31), which satisfies

- (i) $\max_k x_{t_k} - \varepsilon \log N \leq \text{SM}_\varepsilon^{(h)}(x.) \leq \max_k x_{t_k}$;
- (ii) the 1-Lipschitz property on the grid: $\left| \text{SM}_\varepsilon^{(h)}(x.) - \text{SM}_\varepsilon^{(h)}(y.) \right| \leq \max_k |x_{t_k} - y_{t_k}|$.

Consequently, the *smoothing bias* obeys, using the exponential moment bounds from Lemma 3.16 and Doob's L^p inequality (as in the proof of Proposition 3.20, which yields $\sup_h \mathbb{E}[e^{M_h}] + \sup_h \mathbb{E}[e^{\widehat{M}_h}] < \infty$),

$$\begin{aligned} & \left| \mathbb{E} \left[(e^{M_h} - K)^+ \right] - \mathbb{E} \left[\left(e^{\text{SM}_\varepsilon^{(h)}(X^\tau) + \varepsilon \log N} - K \right)^+ \right] \right| \\ & + \left| \mathbb{E} \left[(e^{\widehat{M}_h} - K)^+ \right] - \mathbb{E} \left[\left(e^{\text{SM}_\varepsilon^{(h)}(\widehat{X}^\tau) + \varepsilon \log N} - K \right)^+ \right] \right| \leq C_T \varepsilon \log \frac{T}{h}. \end{aligned} \quad (\text{B.40})$$

Set

$$A_2 := \left| \mathbb{E} \left[\left(e^{\text{SM}_\varepsilon^{(h)}(\widehat{X}^\tau) + \varepsilon \log N} - K \right)^+ \right] - \mathbb{E} \left[\left(e^{\text{SM}_\varepsilon^{(h)}(X^\tau) + \varepsilon \log N} - K \right)^+ \right] \right|.$$

By $|e^u - e^v| \leq (e^u + e^v)|u - v|$ and the grid Lipschitz property,

$$A_2 \leq C_T \mathbb{E} \left[\max_{1 \leq k \leq N} \left| \widehat{X}_{t_k}^\tau - X_{t_k}^\tau \right| \right] \leq C_T \left(\sum_{k=1}^N \mathbb{E} |\widehat{X}_{t_k}^\tau - X_{t_k}^\tau|^2 \right)^{1/2}.$$

Let $\Delta X_t := \widehat{X}_t^\tau - X_t^\tau$ and $\Delta M_t := (M_t^{\tau,X} - W_t^X, M_t^{\tau,V} - W_t^V)$. By the “driver-to-state” L^2 stability in Theorem 3.10, for each $t \leq T$,

$$(\mathbb{E} |\Delta X_t|^2)^{1/2} \leq C_T \mathfrak{d}_\tau, \quad \mathfrak{d}_\tau^2 := \int_0^T \mathbb{E} |\Delta M_s|^2 ds.$$

Hence

$$\frac{1}{N} \sum_{k=1}^N \mathbb{E} |\Delta X_{t_k}|^2 \leq \sup_{t \leq T} \mathbb{E} |\Delta X_t|^2 \leq C_T \mathfrak{d}_\tau^2,$$

and therefore

$$A_2 \leq C_T N^{1/2} \mathfrak{d}_\tau \asymp C_T h^{-1/2} \mathfrak{d}_\tau. \quad (\text{B.41})$$

By Proposition 3.9, $\mathfrak{d}_\tau^2 = \int_0^T \mathbb{E} |M_s^\tau - W_s|^2 ds \leq C_T \tau^{-1/2}$, so

$$A_2 \leq C_T h^{-1/2} \tau^{-1/4}. \quad (\text{B.42})$$

Collecting (B.39), (B.40), and (B.42), we get

$$(i) \leq C_{T,K} \left(\sqrt{h \log(T/h)} + \varepsilon \log \frac{T}{h} + h^{-1/2} \tau^{-1/4} \right).$$

Choose $\varepsilon := h$ and set $h = \tau^{-1/4}$. Since $T \leq 1$ is fixed, the smoothing bias $h \log(T/h) = \tau^{-1/4} \log(T\tau^{1/4})$ is asymptotically smaller than the other terms. The coupling term is $h^{-1/2} \tau^{-1/4} = \tau^{-1/8}$. The discretization term is $\sqrt{h \log(T/h)} \asymp \tau^{-1/8} \sqrt{\log \tau}$. The dominant term is thus

$$(i) \leq C_{T,K} \tau^{-1/8} \sqrt{\log \tau}.$$

Therefore,

$$|LB^{\text{INAR},\tau}(T, K) - LB(T, K)| \leq C_{T,K} \left(\tau^{\frac{1}{4} - \frac{\alpha}{2}} + \tau^{-1/8} \sqrt{\log \tau} \right),$$

and the dominance statement follows by comparing exponents.