Received: 1 December 2018 | Accepted: 1 July 2019

DOI: 10.1111/mafi.12233

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics WILEY

**ORIGINAL ARTICLE**

# A regularity structure for rough volatility

**Christian Bayer**[1] | **Peter K. Friz**[2] | **Paul Gassiat**[3] | **Jorg Martin**[4] | **Benjamin Stemper**[2]

[1]WIAS, Berlin, Germany

[2]TU and WIAS, Berlin, Germany

[3]Paris Dauphine University, Paris, France

[4]HU, Berlin, Germany

**Correspondence**
Peter K. Friz, TU and WIAS Berlin, Berlin 10623, Germany.
Email: friz@math.tu-berlin.de

**Funding information**
DFG research grants BA5484/1 and (CB, BS), FR2943/2 (PKF, BS). ERC via Grant CoG-683164 (PKF). ANR via Grant ANR-16-CE40-0020-01 (PG). DFG Research Training Group RTG 1845 (JM).

**Abstract**

A new paradigm has emerged recently in financial modeling: rough (stochastic) volatility. First observed by Gatheral et al. in high-frequency data, subsequently derived within market microstructure models, rough volatility captures parsimoniously key-stylized facts of the entire implied volatility surface, including extreme skews (as observed earlier by Alòs et al.) that were thought to be outside the scope of stochastic volatility models. On the mathematical side, Markovianity and, partially, semimartingality are lost. In this paper, we show that Hairer's regularity structures, a major extension of rough path theory, which caused a revolution in the field of stochastic partial differential equations, also provide a new and powerful tool to analyze rough volatility models.

*Dedicated to Professor Jim Gatheral on the occasion of his 60th birthday.*

## 1 | INTRODUCTION

We are interested in stochastic volatility (SV) models given in Itô differential form

$$dS_t/S_t = \sigma_t dB_t \equiv \sqrt{v_t} dB_t. \tag{1}$$

Here, $B$ is a standard Brownian motion and $\sigma$ (respectively, $v$) is the *stochastic volatility* (respectively, *variance*) process. Many classical Markovian asset price models fall in this framework, including

Dupire's local volatility model, the SABR, Stein–Stein, and Heston models. In all named SV models, one has Markovian dynamics for the variance process

$$dv_t = g(v_t)dW_t + h(v_t)dt. \tag{2}$$

Constant correlation $\rho := d\langle B, W\rangle_t/dt$ is incorporated by working with a 2D standard Brownian motion $(W, \overline{W})$,

$$B := \rho W + \overline{\rho}\overline{W}, \overline{\rho} := \sqrt{1-\rho^2}.$$

This paper is concerned with an important class of non-Markovian (fractional) SV models, dubbed *rough volatility (RV) models*, in which case $\sigma_t$ (equivalently: $v_t \equiv \sigma_t^2$) is modeled via a fractional Brownian motion (fBm) in the regime $H \in (0, 1/2)$. The term "rough" stems from the fact that in such models, SV (variance) sample paths are $(H-\kappa)$-Hölder continuous, for any $\kappa > 0$, hence "rougher" than Brownian sample paths. Note the stark contrast to the idea of "trending" fractional volatility, which amounts to taking $H > 1/2$. The evidence for the rough regime (recent calibration suggests $H$ as low as 0.05) is now overwhelming—both under the physical and the pricing measure (see Alòs, León, & Vives, 2007; Bayer, Friz, & Gatheral, 2016; Forde & Zhang, 2017; Fukasawa, 2011, 2017; Gatheral, Jaisson, & Rosenbaum, 2018; Mijatović & Tankov, 2016). It should be noted, however, that these different regimes can be easily mixed, so that rough volatility governs the short time behavior, while trending volatility affects the long time behavior; we refer to Comte and Renault (1998), Comte, Coutin, and Renault (2012), Alòs and Yang (2017), and Bennedsen, Lunde, and Pakkanen (2016) for more information on this.

Much attention in the above references on rough volatility models has, in fact, been given to "simple" rough volatility models of the form

$$\sigma_t = f(\widehat{W}_t), \tag{3}$$

$$\widehat{W}_t := \int_0^t K(s,t)dW_s, \tag{4}$$

$$K(s,t) := \sqrt{2H}|t-s|^{H-1/2}\mathbf{1}_{t>s}, \quad H \in (0, 1/2). \tag{5}$$

(Later on, we will allow for explicit time dependence of $f$ in order to cover the rough Bergomi model (Bayer et al., 2016).) In other words, volatility is an explicit function of an fBm, with fixed Hurst parameter. More specifically, following Bayer et al. (2016), we work with the Volterra fBm, a.k.a. Riemann–Liouville fBm, but other choices such as the Mandelbrot van Ness fBm, with suitably modified kernel $K$, are possible. Note that, in contrast to many classical SV models (such as Heston), the SV is explicitly given, and no rough or stochastic differential equation needs to be solved (hence the term "simple"). Rough volatility not only provides remarkable fits to both time series and option price data, but it also has a market microstructure justification: starting with a Hawkes process model. Rosenbaum and coworkers (El Euch, Fukasawa, & Rosenbaum, 2018; El Euch & Rosenbaum, 2018, 2019) find, in a suitable scaling limit, functions $f(\cdot), v(\cdot), u(\cdot)$ such that

$$\sigma_t := f(\widehat{Z}_t) \quad \cdots \quad \text{"nonsimple rough volatility (RV)"} \tag{6}$$

$$Z_t = z + \int_0^t K(s,t)v(Z_s)ds + \int_0^t K(s,t)u(Z_s)dW_s. \tag{7}$$

Such stochastic Volterra dynamics provide a natural generalization of simple rough volatility. We refer to this class of models as "nonsimple": in contrast to the aforementioned simple model, (7) generally does not admit a closed-form solution.

## 1.1 | Markovian stochastic volatility models

For comparison with rough volatility, which will be discussed in more detail below, we first mention a selection of tools and methods well known for *Markovian* SV models.

- PDE methods are ubiquitous in (low-dimensional) pricing problems, as are;
- Monte Carlo methods, noting that knowledge of strong and weak rates of convergence of time discretizations of stochastic differential equations (typically with rates $1/2$ and 1, respectively) is the starting point of modern multilevel methods (multilevel Monte Carlo [MLMC]);
- Quasi-Monte Carlo (QMC) methods are widely used; related in spirit we have the Kusuoka–Lyons–Victoir cubature approach, popularized in the form of the Ninomiya–Victoir splitting scheme, nowadays available in standard software packages;
- Freidlin–Wentzell's theory of small noise large deviations is essentially immediately applicable, as are various "strong" large deviations (a.k.a. exact asymptotic) results, used, for example, to derive the famous SABR formula.

For several reasons, it can be useful to write model dynamics in *Stratonovich form*: From a PDE perspective, the operators then take a sum-of-squares form that can be exploited in many ways (think Hörmander theory, Malliavin calculus, etc.). From a numerical perspective, we note that the Kusuoka–Lyons–Victoir scheme (Kusuoka, 2001; Lyons & Victoir, 2004) also requires the full dynamics to be rewritten in Stratonovich form. In fact, viewing the Ninomiya–Victoir scheme Ninomiya and Victoir (2008) as level-5 cubature, in the sense of Lyons and Victoir (2004), its level-3 variant is nothing but the familiar Wong–Zakai approximation for diffusions. Another financial example that requires a Stratonovich formulation comes from interest rate model validation (Davis & Mataix-Pastor, 2007), based on the Stroock–Varadhan support theorem. We further note that QMC (based on Sobol numbers, say) works particularly well if the noise has a multiscale decomposition, as obtained by interpreting a (piecewise) linear Wong–Zakai approximation as a Haar wavelet expansion of the driving white noise. Indeed, the naturally induced order of random coefficients, in terms of their importance, leads to a lower "effective dimension" of the integration problem, see, for instance, Acworth, Broadie, and Glasserman (1998).

## 1.2 | Complications with rough volatility

Due to loss of Markovianity, PDE methods are not applicable, and nor are (off-the-shelf) Freidlin–Wentzell large deviation estimates (but see Forde & Zhang, 2017). Moreover, the variance process in rough volatility models is not a semimartingale, which complicates the use of several established stochastic analysis tools. In particular, *rough volatility admits no Stratonovich formulation*. Closely related, one lacks a (Wong–Zakai type of) approximation theory for rough volatility. To see this, focus on the "simple" situation, that is, (1) and (3), so that

$$S_t/S_0 = \mathcal{E}\left(\int_0^\cdot f\left(\widehat{W}_s\right) dB_s\right)(t). \tag{8}$$

Inside the (classical) stochastic exponential $\mathcal{E}(M)(t) = \exp(M_t - \frac{1}{2}[M]_t)$ we have the martingale term

$$\int_0^t f(\widehat{W}_s) dB_s = \rho \underbrace{\int_0^t f(\widehat{W}_s) dW_s} + \overline{\rho} \int_0^t f(\widehat{W}_s) d\overline{W}_s. \tag{9}$$

In essence, the trouble is due to the underbraced, innocent looking Itô-integral. Indeed, any naive attempt to put it in Stratonovich form

$$\int_0^t f(\widehat{W}_s)\circ dW_s := \int_0^t f(\widehat{W}_s)dW_s + (\text{Itô} - \text{Stratonovich correction}), \tag{10}$$

or, in the spirit of Wong–Zakai approximations,

$$\int_0^t f(\widehat{W}_s)\circ dW_s := \lim_{\varepsilon\to 0}\int_0^t f(\widehat{W}_s^\varepsilon)dW_s^\varepsilon, \tag{11}$$

must fail for $H < 1/2$. The Itô–Stratonovich correction is given by the quadratic covariation, defined (whenever possible) as the limit, in probability, of

$$\sum_{[u,v]\in\pi}(f(\widehat{W}_v) - f(\widehat{W}_u))(W_v - W_u), \tag{12}$$

along any sequence $(\pi^n)$ of partitions with mesh-size tending to zero. Disregarding trivial situations, this limit does not exist. For instance, when $f(x) = x$, fractional scaling immediately gives divergence (at rate $H - 1/2$) of the expression (12). This issue also arises in the context of option pricing; compare Theorem 1.4 and Section 6 below. All these problems remain present, of course, for the more complicated situation of "nonsimple" rough volatility, as discussed in Section 5.

## 1.3 | Description of main results

Motivated by singular SPDE theory, such as Hairer's work on Kardar–Parisi–Zhang (KPZ) (Hairer, 2013) and the Hairer–Pardoux "renormalized" Wong–Zakai theorem (Hairer & Pardoux, 2015), we provide a (necessarily renormalized) strong approximation theory for rough volatility. Rough path theory, despite its very purpose to deal with low regularity paths, is not applicable to the problem at hand. (We shall elaborate on this at the beginning of Section 2.) In essence, what one needs is a more flexible type of rough paths theory, which is exactly what Hairer's theory of regularity structures Hairer (2014) supplies. As a consequence of fundamental continuity statements in "model" (think: "rough path") metrics, we will discuss short-time large deviations for rough volatility models. Following, for example P. K. Friz and Hairer (2014, Section 9.3), we also envision support results in "rough" interest rate models in the spirit of Davis and Mataix-Pastor (2007).

To state our basic approximation results, write $\dot{W}^\varepsilon := \partial_t W^\varepsilon$ for a suitable approximation at scale $\varepsilon$ to white noise, with the induced approximation to fBm denoted by $\widehat{W}^\varepsilon$. Throughout, the Hurst parameter $H \in (0, 1/2]$ is fixed and $f$ is a smooth function, such that (8) is a (local) martingale, as required by standard financial theory. More precisely, let $W^\epsilon$ denote the Haar wavelet construction of the Brownian motion $W$ truncated at level $N = -\log_2(\epsilon)$, see Section 3.4 for details. Then, $\dot{W}^\epsilon$ is simply defined as the time derivative of the (piecewise linear) process $W^\epsilon$ and $\widehat{W}^\epsilon$ is obtained by integrating (in a pathwise fashion) $\dot{W}^\epsilon$ against the Volterra kernel $K$, that is, $\widehat{W}_t^\epsilon = \int_0^t K(t,s)\dot{W}_s^\epsilon ds$. $B^\epsilon$ denotes the analogous construction for a correlated Brownian motion $B$, where $B_t = \rho W_t + (1-\rho^2)^{1/2}\overline{W}_t$, with some independent Brownian motion $\overline{W}$.

**Theorem 1.1.** *Consider a simple rough volatility model with dynamics $dS_t/S_t = f(\widehat{W}_t)dB_t$, that is, driven by Brownian motions $B$ and $W$ with constant correlation $\rho$. Then, there exist $\varepsilon$-periodic* *functions* $\mathscr{C}^\varepsilon = \mathscr{C}^\varepsilon(t)$, *with diverging averages* $C_\varepsilon$, *such that a Wong–Zakai result holds of the form that* $\widetilde{S}^\varepsilon \to S$ *in probability and uniformly on compacts, where*

$$\frac{\partial_t \widetilde{S}_t^\varepsilon}{\widetilde{S}_t^\varepsilon} = f(\widehat{W}_t^\varepsilon)\dot{B}_t^\varepsilon - \rho \mathscr{C}^\varepsilon(t) f'(\widehat{W}_t^\varepsilon) - \tfrac{1}{2} f^2(\widehat{W}_t^\varepsilon), \quad \widetilde{S}_0^\varepsilon = S_0.$$

*Similar results hold for more general ("nonsimple") rough volatility models.*

*Remark* 1.2. When $H = 1/2$, this result is an easy consequence of the well-known Itô–Stratonovich conversion formula. In the case $H < 1/2$, Theorem 1.1 provides the interesting insight that genuine renormalization (in the sense of subtracting diverging quantities) is required *if and only* if the correlation parameter $\rho$ is nonzero. This is the case in equity (and many other) markets. Also note that naive approximations without renormalization (i.e., without subtracting the $\mathscr{C}^\varepsilon$-term) will in general diverge.

*Remark* 1.3. Mollification of the noise by truncation of the wavelet representation of the driving Brownian motion is natural for numerical purposes. First, it gives a simple sampling technique in terms of independent, identically distributed (IID) standard normals. Second, the construction provides a canonical hierarchy that is beneficial for QMC methods, compare the discussion in Section 1.1.

To formulate implications for option pricing, define the Black–Scholes pricing function

$$C_{BS}\big(S_0, K; \sigma^2 T\big) := \mathbb{E}\left(S_0 \exp\left(\sigma\sqrt{T}Z - \frac{\sigma^2}{2}T\right) - K\right)^+, \tag{13}$$

where $Z$ denotes a standard normal random variable. We then have following theorem.

**Theorem 1.4.** *With* $\mathscr{C}^\varepsilon = \mathscr{C}^\varepsilon(t)$ *as in Theorem 1.1, define the renormalized integral approximation*

$$\widetilde{\mathscr{I}}^\varepsilon := \widetilde{\mathscr{I}}_f^\varepsilon(T) := \int_0^T f(\widehat{W}_t^\varepsilon) dW_t^\varepsilon - \int_0^T \mathscr{C}^\varepsilon(t) f'(\widehat{W}_t^\varepsilon) dt, \tag{14}$$

*and also the approximate total variance*

$$\mathscr{V}^\varepsilon := \mathscr{V}_f^\varepsilon(T) := \int_0^T f^2(\widehat{W}_t^\varepsilon) dt.$$

*Then the price of a European call option, under the pricing model (1), (3), struck at* $K$ *with time* $T$ *to maturity, is given by*

$$\mathbb{E}\Big[\big(S_T - K\big)^+\Big] = \lim_{\varepsilon \to 0} \mathbb{E}\Big[\Psi(\widetilde{\mathscr{I}}^\varepsilon, \mathscr{V}^\varepsilon)\Big],$$

*where*

$$\Psi(\mathscr{I}, \mathscr{V}) := C_{BS}\left(S_0 \exp\left(\rho \mathscr{I} - \frac{\rho^2}{2}\mathscr{V}\right), K, \bar{\rho}^2 \mathscr{V}\right). \tag{15}$$

*Similar results hold for more general ("nonsimple") rough volatility models.*

Let us discuss right away how to reduce the statements of Theorems 1.1 and 1.4 to the actual convergence statements that will occupy us in Section 3 of the main text. First, note that

$$S_t = S_0 \exp\left[\int_0^t f\Big(\widehat{W}_s\Big) dB_s - \tfrac{1}{2}\int_0^t f^2\Big(\widehat{W}_s\Big) ds\right]. \tag{16}$$

The approximations $W^\varepsilon, \overline{W}^\varepsilon$, and $B^\varepsilon := \rho W^\varepsilon + \overline{\rho}\overline{W}^\varepsilon$ converge uniformly to the obvious limits, so that it suffices to understand the convergence of the stochastic integral. Note that $\widehat{W}$ is heavily correlated with $W$ but independent of $\overline{W}$. The interesting part is then the convergence

$$\int_0^t f(\widehat{W}_s^\varepsilon)dW_s^\varepsilon - \int_0^t \mathscr{C}^\varepsilon(s) f'(\widehat{W}_s^\varepsilon)ds \to \int_0^t f(\widehat{W}_s)dW_s, \tag{17}$$

as stated and proved in Theorem 3.25. For the other part, no correction terms arise due to independence, and it can be seen with standard methods that

$$\int_0^t f(\widehat{W}^\varepsilon)_s d\overline{W}_s^\varepsilon \to \int_0^t f(\widehat{W}_s)d\overline{W}_s,$$

in the sense of convergence in probability, uniformly on compacts in $t$. The convergence result of Theorem 1.1 then follows readily. As for pricing, in Theorem 1.4, we consider the call payoff

$$\left(S_0 \exp\left[\int_0^T \sigma_t dB_t - \frac{1}{2}\int_0^T \sigma_t^2 dt\right] - K\right)^+.$$

An elementary conditioning argument w.r.t. $W$ (first used by Romano–Touzi in the context of Markovian SV models), and then shows that the call price is given as expectation of

$$C_{BS}\left(S_0 \exp\left(\rho\int_0^T \sigma_t dW_t - \frac{\rho^2}{2}\int_0^T \sigma_t^2 dt\right), K, \frac{\overline{\rho}^2}{2}\int_0^T \sigma_t^2 dt\right).$$

Specializing to the case $\sigma_t = f(\widehat{W}_t)$, in combination with Theorem 3.25, then yields Theorem 1.4. Note that extensions to nonsimple RV are immediate from suitable extensions of Theorem 3.25, as discussed in Section 5.2.

From a *mathematical perspective*, the key issue in proving the above theorems is to establish convergence of the *renormalized* approximate integrals, as $\varepsilon \to 0$,

$$\widetilde{\mathscr{I}}^\varepsilon = \int_0^T f(\widehat{W}_t^\varepsilon)dW_t^\varepsilon - \int_0^T \mathscr{C}^\varepsilon(t) f'(\widehat{W}_t^\varepsilon)dt \to \int_0^T f(\widehat{W}_t)dW_t. \tag{18}$$

Here, we find much inspiration from singular SPDE theory, which also requires renormalized approximations for convergence to the correct Itô-object. Specifically, we see that the theory of regularity structures (Hairer, 2014), which essentially emerged from the theory of rough paths and Hairer's KPZ analysis (see P. K. Friz & Hairer, 2014, for a discussion and references), is a very appropriate tool here. In turn, we add an interesting new class of examples to the existing instances of regularity structures (polynomials, rough paths, many singular SPDEs, etc.). This new example avoids all considerations related to spatial structure (notably multilevel Schauder estimates; cf. Hairer, 2014, Chapter 5), yet comes with the genuine need for renormalization. In fact, because we do not restrict ourselves to approximations of the white noise obtained by mollification (i.e., by convolution of $\dot{W}$ with a rescaled mollifier function, say $\delta^\varepsilon(x,y) = \varepsilon^{-1}\rho(\varepsilon^{-1}(y-x))$), our analysis naturally leads us to *renormalization functions*. In case of mollifier approximations, which is the usual choice of Hairer and coworkers (Chandra & Hairer, 2016; Hairer, 2013, 2014)—but rules out wavelet approximations—the renormalization function turns out to be constant (because $\dot{W}^\varepsilon$ is still stationary). In this case, we would obtain

$$\mathscr{C}^\varepsilon(t) \equiv C_\varepsilon = c\varepsilon^{H-1/2},$$

with $c = c(\rho)$ explicitly given by an integral, compare (40). If, on the other hand, we consider a Haar wavelet approximation of white noise, we get[1]

$$\mathscr{C}^{\varepsilon}(t) = \frac{\sqrt{2H}}{H+1/2}\frac{|t-\lfloor t/\varepsilon\rfloor\varepsilon|^{H+1/2}}{\varepsilon} \quad \text{with mean } C_{\varepsilon} = \frac{\sqrt{2H}}{(H+1/2)(H+3/2)}\varepsilon^{H-1/2}. \tag{19}$$

It is natural to ask if $\mathscr{C}^{\varepsilon}(t)$ can be replaced, after all, by its mean $C_{\varepsilon}$. (Of course, this mean is still diverging for $\varepsilon \to 0$, as $H < 1/2$.) For $H > 1/4$, the answer is yes, with an interesting phase transition when $H = 1/4$, compare Section 3.2.

From a *numerical perspective*, Theorem 1.4 avoids any sampling of the independent factor $\overline{W}$. A brute force approach then consists in simulating a scalar Brownian motion $W$, followed by discrete approximation of the stochastic integral $= \widehat{W}_t = \int_0^t K(t,s)dW_s$. However, given the singularity of the Volterra kernel $K$, this is not advisable and it is preferable to simulate the two-dimensional Gaussian process $(W_t, \widehat{W}_t : 0 \le t \le T)$, whose covariance function is readily available. A remaining problem is that the speed of convergence of

$$\sum f(\widehat{W}_s)W_{s,t} \to \int_0^T f\left(\widehat{W}_t\right)dW_t,$$

with $[s,t]$ taken in a partition of mesh size $\sim 1/n$, is very slow as $\widehat{W}$ has little regularity when $H$ is small (Bayer et al., 2016; Gatheral et al., 2018, report $H \approx 0.05$.). Here, higher order approximations come to help, and we include quantitative estimates, more precisely: *strong* rates, throughout. Such rates are essential for the design of MLMC algorithms, as was also seen in the context of general Gaussian rough differential equations (Bayer, Friz, Riedel, & Schoenmakers, 2016). (The important analysis of *weak rates* is left for future work.) Numerical aspects are further discussed in Section 6.

The second set of results concerns large deviations for rough volatility models. Thanks to the contraction principle and fundamental continuity properties of Hairer's reconstruction map, the problem is reduced to understanding a large deviations principle (LDP) for a suitable enhancement of the noise. This approach requires (sufficiently) smooth coefficients, but comes with no growth restrictions, which is indeed quite suitable for financial modeling: we improve the Forde–Zhang short-time large deviations (Forde & Zhang, 2017) for simple rough volatility models such as to include $f$ of exponential type, a defining feature in the works of Gatheral and coauthors (Bayer et al., 2016; Gatheral et al., 2018). (Such an extension is also subject of the recent works Jacquier, Pakkanen, & Stone, 2018; Gulisashvili, 2018.)

**Theorem 1.5.** *Let $X_t = log(S_t/S_0)$ be the log-price under simple rough SV model, that is, (1) and (3). Then, $(t^{H-\frac{1}{2}}X_t : t \ge 0)$ satisfies a short-time large-deviation principle (LDP) with speed $t^{2H}$ and rate function given by*

$$I(y) = \inf_{h\in L^2([0,1])}\left\{\frac{1}{2}\|h\|_{L^2}^2 + \frac{\left(y-\rho I_1(h)\right)^2}{2I_2(h)}\right\}, \tag{20}$$

*with $I_1(h) := \int_0^1 f(\widehat{h}(t))h(t)dt$, $I_2(h) := \int_0^1 f(\widehat{h}(t))^2dt$, where $\widehat{h}(t) := \int_0^t K(s,t)h(s)ds$.*

Theorem 1.5 is proved below as Corollary 4.3.

*Remark* 1.6. A potential short coming is the nonexplicit form of the rate function. Geometric or "Hamiltonian" interpretations of the rate function, studied in a Markovian setting by many authors (e.g., Avellaneda, Boyer-Olson, Busca, & Friz, 2003; Bayer & Laurence, 2014; Berestycki, Busca, & Florent, 2004; Deuschel et al., 2014a; Deuschel, Friz, Jacquier, & Violante, 2014b), are then lost. A partial remedy here is to move from large deviations to (higher order) *moderate deviations*. Analytic tractability is so restored and one still captures the main feature of the volatility smile close to the money. This method was introduced in a Markovian setting in P. Friz, Gerhold, and Pinter (2018), the extension to simple rough volatility models was given in Bayer, Friz, Gulisashvili, Horvath, and Stemper (2019), relying either on Forde and Zhang (2017) or the above Theorem 1.5.

We next turn to nonsimple rough volatility models. Inspired by Rosenbaum and coworkers (El Euch et al., 2018; El Euch & Rosenbaum, 2019, 2018), we consider the stochastic Itô–Volterra equation

$$Z_t = z + \int_0^t K(s,t)\big(u(Z_s)dWs + v(Z_s)ds\big),$$

with corresponding log-price process given by

$$X_t = \int_0^t f(Z_s)(\rho dW_s + \overline{\rho} d\overline{W}_s) - \frac{1}{2}\int_0^t f^2(Z_s)ds.$$

(For simplicity, we here consider $f, u, v$ to be bounded, with bounded derivatives of all orders.) For $h \in L^2([0,T])$, let $z^h$ be the unique solution to the integral equation

$$z^h(t) = z + \int_0^t K(s,t)u(z^h(s))h(s)ds,$$

and define $I_1^z(h) := \int_0^1 f(z^h(s))h(s)ds$ and $I_2^z(h) := \int_0^1 f(z^h(s))^2 ds$. Then, we have the following extension of Theorem 1.5 (and also Forde & Zhang, 2017; Gulisashvili, 2018; Jacquier et al., 2018) to nonsimple rough volatility.

**Theorem 1.7.** *Let $X_t := log(S_t/S_0)$ be the log-price under nonsimple rough SV and assume $H > 1/4$. Then, $t^{H-\frac{1}{2}}X_t$ satisfies an LDP with speed $t^{2H}$ and rate function given by*

$$I(x) = \inf_{h \in L^2([0,T])}\left\{\frac{1}{2}\|h\|_{L^2}^2 + \frac{\big(x - \rho I_1^z(h)\big)^2}{2I_2^z(h)}\right\}. \tag{21}$$

Theorem 1.7 is proved below as Corollary 5.5.

*Remark* 1.8. We showed in (Bayer et al., 2019, Corollary 11)—but see related results by Alòs et al. Alòs et al. (2007) and Fukasawa (Fukasawa, 2011, 2017)—that in the previously considered simple rough volatility models, now writing $\sigma(.)$ instead of $f(.)$, the *implied volatility skew* behaves, in the short time limit, as $\sim \rho\frac{\sigma'(0)}{\sigma(0)}\langle K1, 1\rangle t^{H-1/2}$, where $\langle K1,1\rangle$ in our setting computes to $c_H := \frac{(2H)^{1/2}}{(H+1/2)(H+3/2)}$. (The blowup $t^{H-1/2}$ as $t \to 0$ is a desired feature, in agreement with steep skews seen in the market.) To first order, $Z_t \approx z + u(z)\int_0^t K(s,t)dWs = z + u(z)\widehat{W} =: \sigma(\widehat{W})$, from which one obtains a skew formula in the nonsimple rough volatility case of the form

$$\rho u(z)\frac{f'(z)}{f(z)}c_H t^{H-1/2}.$$

Following the approach of Bayer et al. (2019), Theorem 1.7 not only allows for rigorous justification of these formulas, but also for the computation of higher order smile features, although this is not pursued in this article. In the case of classical (Markovian) SV models, $H = 1/2$, and specializing further to $f(x) \equiv x$, so that $Z$ (respectively, $z$) models stochastic (respectively, spot) volatility, this formula reduces precisely to the popular skew formula from Gatheral's book (Gatheral & Taleb, 2006, (7.6)), attributed therein to Medvedev–Scaillet. In the case of the *rough Heston* model, where $Z$ models stochastic variance, compare (54), we have $f(x) = \sqrt{x}$, $u = \eta\sqrt{x}$ and this leads to the following short-dated skew formula:

$$\frac{\rho\eta}{2\sqrt{v_0}} c_H t^{H-1/2}.$$

If the above expression is multiplied with $2\sqrt{v_0}$, we get the implied variance skew, again in agreement with Gatheral (Gatheral & Taleb, 2006, p. 35). The formula may be independently verified via the characteristic function obtained in El Euch and Rosenbaum (2019).

The reader may be interested in further applications of the regularity structure view on rough volatility developed in this paper. The Stratonovich formulation opens up the possibility of constructing cubature methods (in the sense of Kusuoka, 2001; Lyons & Victoir, 2004) for rough SV methods. Indeed, our method can be seen as a level-3 Ninomiya-Victoir Ninomiya and Victoir (2008) scheme. Further, having said much about large deviations, it is not far-fetched to think about a support theorem (another classical application area of rough paths and regularity structures, cf. P. K. Friz & Hairer, 2014, Section 9.3), which, in turn, invites to revisit Davis and Mataix-Pastor (2007) in a setting of "rough" interest rate models. Another concrete application, content of the recent P. K. Friz, Gassiat, and Pigato (2018), concerns precise asymptotics, allowing for considerable refinement of large deviations. (Translated to financial terms, this improvement leads to higher order implied volatility expansions.)

**Structure of the article.** In Section 2, we explain why the classical formulation of rough paths is not suitable for rough volatility models, and then go on to introduce essentials of the theory of regularity structures. We use the KPZ equation as a guiding example, which offers several similarities to rough volatility. The most basic "pricing structure" is introduced in Section 3. In Section 4, we consider a regularity structure for two-dimensional noise, which is necessary to study the asset price process in addition to the volatility process. Section 5 then discusses the case of nontrivial dynamics for rough volatility. Some numerical results are presented in Section 6, followed by several appendices with technical details. From Section 3 on, all our work relies on the framework of Hairer's regularity structures. There seems to be no point in repeating all the necessary definitions and terminology, which the reader can find in Hairer (2013), Hairer (2014), Hairer (2015), and P. K. Friz and Hairer (2014) and a variety of survey papers on the subject. (For the reader in search of one concise reference, we recommend P. K. Friz & Hairer, 2014, Section 13.)

Participants of Global Derivatives 2017 (Barcelona) and Gatheral 60th Birthday conference (CIMS, NYU) are thanked for their valuable feedback. We are also very thankful to anonymous referees for their very constructive feedback.

## 2 | ON ROUGH PATHS AND LESSONS FROM KPZ AND SINGULAR SPDE THEORY

We already pointed out in Section 1.2 that any analysis of *correlated* ($\rho \neq 0$) rough volatility models will involve the (Itô-) integral

$$\int_0^T f(\widehat{W}_t) dW_t, \tag{22}$$

where $(\widehat{W}_t)$ is a fBm with Hurst parameter $H < 1/2$, itself given as an integral of a singular kernel against the Brownian motion $W$. Although the scalar Brownian motion $W$ can easily be lifted to a Brownian rough path (of Itô-, or Stratonovich type), the integral in (22) *cannot* be viewed as rough integral. Indeed, a generic integrand $f(\widehat{W})$, and even $\widehat{W}$, is not at all a rough path controlled by $W$ (in the sense of Gubinelli). Hence, the stochastic integral cannot be defined by standard rough path theory as found in P. K. Friz and Hairer (2014, Section 4).

A second attempt, to view $(W, \widehat{W})$ as a two-dimensional Gaussian process—in order to use Gaussian rough path theory (e.g., P. K. Friz & Hairer, 2014, Section 12)—also fails. In fact, this theory would require $W$ and $\widehat{W}$ being independent ($\leftrightarrow$ here: fully correlated), the resulting lift being of geometric (i.e., Stratonovich) type ( $\leftrightarrow$ here: Itô integral), the Hurst parameter $H > 1/4$ ( $\leftrightarrow$ here: any $H > 0$), and an enhancement of at most three levels (level-3 rough paths) ( $\leftrightarrow$ here: $\sim 1/H$ levels).

A third attempt, in view of the nongeometric nature of Itô integration, to use *branched* rough paths is also doomed, for it requires—like classical geometric rough path theory—all iterated integrals. In our case, there is already an obstacle at level-2, before the appearance of any branching, in that the full set of second iterated integrals

$$\begin{pmatrix} \int W dW, & \int W d\widehat{W} \\ \int \widehat{W} dW & \int \widehat{W} d\widehat{W} \end{pmatrix} = \begin{pmatrix} * & ? \\ * & ? \end{pmatrix} \tag{23}$$

is an ill-defined object ($*$ stands for well-defined Itô-integrals, ? for integrals of unclear meaning). Note that imposition of a first-order (respectively, Itô) product rule would manifestly clash with Itô-calculus, as $\widehat{W}$ has infinite quadratic variation when $H < 1/2$.

On the other hand, formal expansion of (22) over some interval $[s, t]$ gives

$$\int_s^t f(\widehat{W}_u) dW_u \approx f\left(\widehat{W}_s\right) W_{s,t} + f'\left(\widehat{W}_s\right) \int_s^t \widehat{W}_{s,u} dW_u + \cdots,$$

so that the troubling terms—the "?" in (23)—do not appear. What is needed then, in the general case, is a higher order "partial" branched rough path theory (for we deal with nongeometric/Itô objects), in which only partial information on the iterated integrals is stored. But even then, one faces failure of canonical (Wong-Zakai type) approximations, that is,

$$\nexists \lim_{\varepsilon \to 0} \int_0^T f(\widehat{W}_t^\varepsilon) dW_t^\varepsilon. \tag{24}$$

Such failures are atypical for rough path theory. Having made all these observations, Hairer's regularity structures (see below for more details) provides everything we desire: a tailor-made algebraic structure (which by construction only stores the required higher order information), together with a machinery that gives continuity properties of all operations of interest *and* a consistent way to renormalize approximate stochastic integrals, such as the one appearing in (24).

The absence of a canonical approximation theory, as seen in (24), is a defining feature of the singular SPDEs recently considered by Hairer, Gubinelli, and now many others. In particular, approximation of the noise (say, $\varepsilon$-mollification for the sake of argument) typically does *not* give rise to convergent approximations. To be specific, it is instructive to review the very example that led Hairer to regularity structures: the universal model for fluctuations of interface growth given by the KPZ equation

$$\partial_t u = \partial_x^2 u + |\partial_x u|^2 + \xi$$

with space-time white noise $\xi = \xi(x, t; \omega)$. As a matter of fact, and without going into further details, there is a well-defined Itô-solution $u = u(t, x; \omega)$ (known as the "Cole-Hopf" solution), but if one considers the equation with $\varepsilon$-mollified noise, then $u = u^\varepsilon$ diverges with $\varepsilon \to 0$. In this sense, there is a fundamental *lack of approximation theory* and *no Stratonovich solution* to KPZ exists. To see the problem, take $u_0 \equiv 0$ for simplicity and write

$$u = H \star \left(|\partial_x u|^2 + \xi\right)$$

with the space-time convolution denoted by $\star$ and the heat kernel

$$H(t, x) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4t}\right) \mathbf{1}_{\{t>0\}}.$$

One can proceed with a Picard iteration

$$u = H \star \xi + H \star ((H' \star \xi)^2) + \cdots,$$

but there is an immediate problem with $(H' \star \xi)^2$, (naively) defined as the $\varepsilon$-to-zero limit of $(H' \star \xi^\varepsilon)^2$, which does not exist. However, there exists a diverging sequence of real numbers $C_\varepsilon$ such that, in probability,

$$\exists \lim_{\varepsilon \to 0} (H' \star \xi^\varepsilon)^2 - C_\varepsilon \rightsquigarrow \text{(new object)} =: (H' \star \xi)^{\diamond 2}.$$

The idea of Hairer, following the philosophy of rough paths, was then to accept

$$H \star \xi, (H' \star \xi)^{\diamond 2} \text{ (and a few more)}$$

as enhancement of the noise ("*model*") upon which solution depends in pathwise robust fashion. This unlocks the seemingly fixed (and here even nonsensical) relation

$$H \star \xi \mapsto \xi \mapsto (H' \star \xi)^2.$$

Loosely speaking, one has

**Theorem 2.1** (Hairer)**.** *There exist diverging constants $C_\varepsilon$ such that a Wong–Zakai*[2] *result holds of the form $\widetilde{u}^\varepsilon \to u$, in probability and uniformly on compacts, where*

$$\partial_t \widetilde{u}^\varepsilon = \partial_x^2 \widetilde{u}^\varepsilon + |\partial_x \widetilde{u}^\varepsilon|^2 - C_\varepsilon + \xi^\varepsilon.$$

*Similar results hold for a number of other singular semilinear SPDEs.*

In a sense, this can be traced back to the *Milstein scheme* for SDEs and then *rough path theory*. Consider $dY_t = f(Y_t) dW_t$, with $Y_0 = 0$ for simplicity, and consider the second-order (Milstein) approximation

$$Y_{t_{i+1}} \approx Y_{t_i} + f\left(Y_{t_i}\right) W_{t_i, t_{i+1}} + f f'\left(Y_{t_i}\right) \int_{t_i}^{t_{i+1}} W_{t_i, s} \dot{W}_s ds.$$

One has to unlock the seemingly fixed relation

$$W \mapsto \dot{W} \mapsto \int_0^{\cdot} W_s \dot{W}_s ds =: \mathbb{W},$$

for there is a choice to be made. For instance, the last term can be understood as Itô-integral $\int_0^{\cdot} W_s dW_s$ or as Stratonovich integral $\int_0^{\cdot} W_s \circ dW_s$ (and, in fact, there are many other choices, see, for example, the discussion in P. K. Friz & Hairer, 2014.) It suffices to take this thought one step further to arrive at *rough path theory*: accept $\mathbb{W}$ as new (analytic) object, which leads to the main (rough path) insight

$$\text{SDE theory} = \text{analysis based on } (W, \mathbb{W}).$$

In comparison,

$$\text{SPDE theory } \textit{à la Hairer} = \text{analysis based on (renormalized) enhanced noise}(\xi, \dots).$$

**Inside Hairer's theory**: [3] As motivation, consider the Taylor expansion (at $x$) of a real-valued smooth function,

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2} f''(x)(y-x)^2 + \cdots .$$

It can be written as an abstract polynomial ("jet") at $x$,

$$F(x) := f(x)\, 1 + g(x) X + h(x) X^2 + \cdots ,$$

with, necessarily, $g = f',\ h = f''/2, \ldots$. If we "realize" these abstract symbols again as honest monomials, that is, $\Pi_x : X^k \mapsto (. - x)^k$, and extend $\Pi_x$ linearly, then we recover the full Taylor expansion:

$$\Pi_x[F(x)](.) = f(x) + g(x)(. - x) + \frac{1}{2} h(x)(. - x)^2 + \cdots$$

Hairer looks for solutions of this form: at every space-time point, a jet is attached, which in case of KPZ turns out—after solving an abstract fixed point problem—to be of the form

$$U(x,s) = u(x,s)\, 1 + \mathsf{I} + \mathsf{Y} + v(x,s)\, X + 2\, \Psi + v(x,s) \angle .$$

As before, every symbol is given concrete meaning by "realizing" it as an honest function (or Schwartz distribution). Naturally, one can take

$$\mathsf{I} \mapsto H \star \xi^{\epsilon} (\text{mollified noise}), \textbf{ or: } \quad \mathsf{I} \mapsto H \star \xi (\text{original noise}). \tag{25}$$

More interestingly, $\mathsf{Y}$ can be mapped to any of

$$\begin{cases} H \star (H' \star \xi^{\epsilon})^2, & \text{canonically enhanced mollified noise; } \textbf{or} \\ H \star [(H' \star \xi^{\epsilon})^2 - C_{\epsilon}], & \text{renormalized} \sim \textbf{or} \\ H \star (H' \star \xi)^{\diamond 2}, & \text{renormalized enhanced noise.} \end{cases} \tag{26}$$

This realization map is called "model" and captures exactly a typical, but otherwise fixed, realization of the noise (mollified or not) together with some enhancement thereof, renormalized or not. For instance, writing $\Pi_{x,s}$ for the realization map for renormalized enhanced noise, one has

$$\Pi_{x,s}[U(x,s)](.) = u(x,s) + H \star \xi|_{(*)} + H \star (H' \star \xi)^{\diamond 2}|_{(*)} + \cdots$$

where $(*)$ indicates suitable centering at $(x, s)$. Mind that $U$ takes values in a (finite) linear space spanned by (sufficiently many) symbols,

$$U(x, s) \in \langle \dots, 1, \mathsf{I}, \curlyvee, X, \curlyvee, <, \dots \rangle =: \mathcal{T}$$

The map $(x, s) \mapsto U(x, s)$ is an example of a *modeled distribution*, the precise definition is a mix of suitable analytic and algebraic conditions (similar to the notation of a controlled rough path).

The analysis requires keeping track of the *degree* (a.k.a. *homogeneity*) of each symbol. For instance, $|\mathsf{I}| = 1/2 - \kappa$ (related to the Hölder regularity of the realized object one has in mind), $|X^2| = 2$, and so on. All these degrees are collected in an *index set*. To compare jets at different points (think $(X - \delta 1)^3 = \cdots$), a group of linear maps on $\mathcal{T}$ is used, called a *structure group*. Last not least, the *reconstruction map* uniquely maps modeled distributions to functions or Schwartz distributions. (This can be seen as generalization of the *sewing lemma*, the essence of rough integration, see, for example, P. K. Friz & Hairer, 2014, which turns a collection of sufficiently compatible local expansions into one function or Schwartz distribution.) In the KPZ context, the (Cole-Hopf or Itô) solution is then indeed obtained as reconstruction of the abstract (modeled distribution) solution $U$.

## 3 | THE ROUGH PRICING REGULARITY STRUCTURE

In this section, we develop the approximation theory for integrals of the type $\int f(\widetilde{W})dW$. In the first part, we present the regularity structure and the associated models we will use. In the second part, we apply the reconstruction theorem from regularity structures to conclude our main result, Theorem 3.25.

### 3.1 | Basic pricing setup

We are given a Hurst parameter $H \in (0, 1/2]$, associated with a fBm (in the Riemann–Liouville sense) $\widehat{W}$, and fix an arbitrary $\kappa \in (0, H)$ and an integer

$$M \geq \max\{m \in \mathbb{N} \mid m \cdot (H - \kappa) - 1/2 - \kappa \leq 0\},$$

so that

$$(M + 1)(H - \kappa) - 1/2 - \kappa > 0\,. \tag{27}$$

At this stage, we can introduce the "level-$(M + 1)$" *model space*

$$\mathcal{T} = \Big\langle \{\Xi, \Xi\mathcal{I}(\Xi), \dots, \Xi\mathcal{I}(\Xi)^M, \mathbf{1}, \mathcal{I}(\Xi), \dots, \mathcal{I}(\Xi)^M\} \Big\rangle\,, \tag{28}$$

where $\langle \dots \rangle$ denotes the vector space generated by the (purely abstract) symbols in $\{\dots\}$. We will sometimes write

$$S = S^{(M)} := \{\Xi, \Xi\mathcal{I}(\Xi), \dots, \Xi\mathcal{I}(\Xi)^M, \mathbf{1}, \mathcal{I}(\Xi), \dots, \mathcal{I}(\Xi)^M\}$$

so that $\mathcal{T} = \mathcal{T}^{(M)} = \bigoplus_{\tau \in S} \mathbb{R}\tau$.

*Remark* 3.1. It is useful here and in the sequel to consider as a sanity check the special case $H = 1/2$ in which case we recover the "level-2" rough path structure as introduced in (P. K. Friz & Hairer, 2014, Chapter 13). More specifically, if we take a Hölder exponent $\alpha := 1/2 - \kappa < 1/2$, we may choose $M = 1$. Then, condition (27) is precisely the familiar condition $\alpha > 1/3$.

The interpretation for the symbols in $S$ is as follows: $\Xi$ should be understood as an abstract representation of the white noise $\xi$ belonging to the Brownian motion $W$, that is, $\xi = \dot{W}$ where the derivative is taken in the distributional sense. Note that because we set $W_x = 0$ for $x \le 0$, we have $\dot{W}(\varphi) = 0$ for $\varphi \in C_c^\infty((-\infty, 0))$. The symbol $\mathcal{I}(\dots)$ has the intuitive meaning of "integration against the Volterra kernel," so that $\mathcal{I}(\Xi)$ represents the integration of the white noise against the Volterra kernel, that is,

$$\sqrt{2H} \int_0^t |t-r|^{H-1/2} \mathrm{d}W_r \,,$$

which is nothing but the fBm $\widehat{W}_t$. Symbols like $\Xi\mathcal{I}(\Xi)^m = \Xi \cdot \mathcal{I}(\Xi) \cdot \dots \cdot \mathcal{I}(\Xi)$ or $\mathcal{I}(\Xi)^m = \mathcal{I}(\Xi) \cdot \dots \cdot \mathcal{I}(\Xi)$ should be read as products between the objects above. These interpretations of the symbols generating $\mathcal{T}$ will be made rigorous by the model $(\Pi, \Gamma)$ in the next subsection. Every symbol in $S$ is assigned a homogeneity, which we define by

$$\begin{aligned}
|\Xi\mathcal{I}(\Xi)^m| &:= -1/2 - \kappa + m(H - \kappa),\ m \ge 0,\\
|\mathcal{I}(\Xi)^m| &:= m(H - \kappa),\ m > 0,\\
|\mathbf{1}| &:= 0\,.
\end{aligned}$$

We collect the homogeneities of elements of $S$ in a set $A := \{|\tau| \mid \tau \in S\}$, whose minimum is $|\Xi| = -1/2 - \kappa$. Note that the homogeneities are multiplicative in the sense that $|\tau\tau'| = |\tau| + |\tau'|$ for $\tau, \tau' \in S$ such that $\tau\tau' = \tau \cdot \tau' \in S$ (with the product defined in the obvious way).

At last, our regularity structure comes with a *structure group* $G$, an (abstract) group of linear operators on the model space $\mathcal{T}$, which should satisfy $\Gamma\tau - \tau = \bigoplus_{\tau' \in S:\, |\tau'| < |\tau|} \mathbb{R}\tau'$ and $\Gamma\mathbf{1} = \mathbf{1}$ for $\tau \in S$ and $\Gamma \in G$. We will choose $G = \{\Gamma_h \mid h \in (\mathbb{R}, +)\}$ given by

$$\Gamma_h \mathbf{1} = \mathbf{1},\ \Gamma_h \Xi = \Xi,\ \Gamma_h \mathcal{I}(\Xi) = \mathcal{I}(\Xi) + h\mathbf{1}\,,$$

and $\Gamma_h(\tau' \cdot \tau) = \Gamma_h \tau' \cdot \Gamma_h \tau$ for $\tau', \tau \in S$ for which $\tau \cdot \tau' \in S$ is defined.

### 3.1.1 | The limiting model $(\Pi, \Gamma)$

Let $W$ be a Brownian motion on $\mathbb{R}_+ := [0, \infty)$ and extend it to all of $\mathbb{R}$ by requiring $W_x = 0$ for $x \le 0$. We will frequently use the notations

$$\int_0^t f(t) \mathrm{d}W_t,\ \int_0^t f(t) \diamond \mathrm{d}W_t, \tag{29}$$

which denote the Itô integral and the Skorokhod integral (which boils down to an Itô integral whenever the integrand is adapted), respectively. For background on Skorokhod integration, we refer to Janson (1997, Section 7.3), and Nualart's ICM lecture (Nualart, 2006) is also highly recommended. Skorokhod integrals have the distinct advantage of avoiding the need of an adapted integrand but coincide with Itô integrals once the integrand is adapted. For a reader unfamiliar with the (beautiful) theory of Skorokhod integration, it should be sufficient to simply think of an Itô integral with possibly non-adapted integrands for the purposes of this article. Whenever we make usage of specific properties of Skorokhod integration, we will make this explicit.

From $W$, we now construct the fBm $\widehat{W}$ in the Riemann–Liouville sense with Hurst index $H \in (0, 1/2]$ as

$$\widehat{W}_t := \dot{W} \star K(t) = \sqrt{2H} \int_0^t |t-r|^{H-1/2} \, \mathrm{d}W_r ,$$

where $K(t) = \sqrt{2H} \, \mathbf{1}_{t>0} \, t^{H-1/2}$ denotes the Volterra kernel. We also write $K(s,t) := K(t-s)$.

To give a meaning to the product terms $\Xi \mathcal{I}(\Xi)^k$, we follow the ideas from rough paths and define an "iterated integral" for $s, t \in \mathbb{R}$, $s \le t$, as

$$\mathbb{W}^m_{s,t} := \int_s^t (\widehat{W}_r - \widehat{W}_s)^m \, \mathrm{d}W_r . \tag{30}$$

$\mathbb{W}^m(s,t)$ satisfies the following modification of Chen's relation.

**Lemma 3.2.** *$\mathbb{W}^m$ as defined in (30) satisfies*

$$\mathbb{W}^m_{s,t} = \mathbb{W}^m_{s,u} + \sum_{l=0}^{m} \binom{m}{l} (\widehat{W}_u - \widehat{W}_s)^l \mathbb{W}^{m-l}_{u,t} \tag{31}$$

*for $s, u, t \in \mathbb{R}$, $s \le u \le t$.*

*Proof.* This is a direct consequence of the binomial theorem. □

We extend the domain of $\mathbb{W}^m$ to all of $\mathbb{R}^2$ by imposing Chen's relation for all $s, u, t \in \mathbb{R}$, that is, we set for $t, s \in \mathbb{R}$, $t \le s$,

$$\mathbb{W}^m_{s,t} := - \sum_{l=0}^{m} \binom{m}{l} (\widehat{W}_t - \widehat{W}_s)^l \mathbb{W}^{m-l}_{t,s}. \tag{32}$$

We are now in the position to define a model $(\Pi, \Gamma)$ that gives a rigorous meaning to the interpretation we gave above for $\Xi, \mathcal{I}(\Xi), \Xi\mathcal{I}(\Xi), \ldots$. Recall that in the theory of regularity structures, a model is a collection of linear maps $\Pi_s : \mathcal{T} \to C^1_c(\mathbb{R})'$, $\Gamma_{st} \in G$ for indices $s, t \in \mathbb{R}$ that satisfy

$$\Pi_t = \Pi_s \Gamma_{st}, \tag{33}$$

$$|\Pi_s \tau(\varphi^\lambda_s)| \lesssim \lambda^{|\tau|} , \tag{34}$$

$$\Gamma_{st}\tau = \tau + \sum_{\tau' \in S : \, |\tau'| < |\tau|} c_{\tau'}(s,t)\tau', \text{ with: } c_{\tau'}(s,t)| \lesssim |s-t|^{|\tau| - |\tau'|} , \tag{35}$$

where the loosely stated bounds in (34) and (35) hide a multiplicative constant, which can be chosen uniformly for $\tau \in S$, any $s, t$ in a compact set and for $\varphi^\lambda_s := \lambda^{-1}\varphi(\lambda^{-1}(\cdot - s))$ with $\lambda \in (0,1]$ and $\varphi \in C^1$ with compact support in the ball $B(0,1)$.

We will work with the following "Itô" model $(\Pi, \Gamma)$, which makes our interpretations of the elements of $S$ more precise. (We will [occasionally] write $(\Pi^{\text{Itô}}, \Gamma^{\text{Itô}})$ to avoid confusion with a generic model, which we also denote by $(\Pi, \Gamma)$.)

$$
\begin{aligned}
&\Pi_s \mathbf{1} = 1, && \Gamma_{ts}\mathbf{1} = \mathbf{1},\\
&\Pi_s \Xi = \dot{W}, && \Gamma_{ts}\Xi = \Xi,\\
&\Pi_s \mathcal{I}(\Xi)^m = \left(\widehat{W}(\cdot) - \widehat{W}(s)\right)^m, && \Gamma_{ts}\mathcal{I}(\Xi) = \mathcal{I}(\Xi) + (\widehat{W}(t) - \widehat{W}(s))\mathbf{1},\\
&\Pi_s \Xi\mathcal{I}(\Xi)^m = \{t \mapsto \tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{W}^m(s,t)\}, && \Gamma_{ts}\tau\tau' = \Gamma_{ts}\tau \cdot \Gamma_{ts}\tau', \quad \text{for } \tau, \tau' \in S \text{ with } \tau\tau' \in S.
\end{aligned}
$$

We extend both maps from $S$ to $\mathcal{T}$ by imposing linearity.

**Lemma 3.3.** *The pair* $(\Pi, \Gamma)$ *as defined above defines (a.s.) a model on* $(\mathcal{T}, A)$.

*Proof.* The only symbol in $S$ for which (33) is not straightforward is $\Xi\mathcal{I}(\Xi)^m$, where the statement follows by Chen's relation. The bounds (34) and (35) follow for $\mathbf{1}$ trivially and for $\mathcal{I}(\Xi)^m$ by the $H-\kappa'$-Hölder regularity of $\widehat{W}$, $\kappa' \in (0, H)$. It is straightforward to check the condition (35) by using the rule $\Gamma_{ts}\tau\tau' = \Gamma_{ts}\tau \cdot \Gamma_{ts}\tau'$ so that we are only left with the task to bound $\Pi_s \Xi\mathcal{I}(\Xi)^m(\varphi_s^\lambda)$. Along the lines of the proof of (P. K. Friz & Hairer, 2014, Theorem 3.1), it follows that $|\mathbb{W}^m_{s,t}| \le C|s-t|^{mH+1/2-(m+1)\kappa}$ (where $C > 0$ denotes a random constant with $C \in \bigcap_{p<\infty} L^p$), so that

$$
\begin{aligned}
|\Pi_s \mathcal{I}(\Xi)^m \Xi(\varphi_s^\lambda)| &= \left| \int \left(\varphi_s^\lambda\right)'(t) \mathbb{W}^m(s,t)\,\mathrm{d}t \right| \le C \int \varphi'^{-1}(t-s))|s-t|^{mH+1/2-(m+1)\kappa} \frac{\mathrm{d}t}{\lambda^2}\\
&\le C\lambda^{mH-1/2-(m+1)\kappa} = C\lambda^{|\mathcal{I}(\Xi)^m\Xi|}.
\end{aligned}
$$

□

As we will see below in Section 3.2, this model is the toolbox from which we can build pathwise Itô integrals of the type $\int_0^t f(r, \widehat{W}(r))\,\mathrm{d}W(r)$. For an approximation theory for such expressions, we are in need of a comparable setup that describes approximations, which will be achieved by introducing a model $(\Pi^\varepsilon, \Gamma^\varepsilon)$.

### 3.1.2 | The approximating model $(\Pi^\varepsilon, \Gamma^\varepsilon)$

The whole definition of the model $(\Pi, \Gamma)$ is based on the object $\dot{W}$. It is therefore natural to build an approximating model by replacing $\dot{W}$ by some modification $\dot{W}^\varepsilon$ that converges (as a distribution) to $\dot{W}$ as $\varepsilon \to 0$.

The definition of $\dot{W}^\varepsilon$ will be based on an object $\delta^\varepsilon$, which should be thought of as an approximation to the Dirac delta distribution. Our goal is to build $\delta^\varepsilon$ from wavelets, which can be as irregular as the Haar functions. We find it therefore convenient to allow $\delta^\varepsilon$ to take values in the Besov space $\mathcal{B}^\beta_{1,\infty}(\mathbb{R})$, $\beta > 1/2 + \kappa$, which includes functions like $\mathbf{1}_{[0,1]} \in \mathcal{B}^1_{1,\infty}(\mathbb{R})$.

*Remark* 3.4. We recall the definition of the Besov space $\mathcal{B}^\beta_{1,\infty}(\mathbb{R})$ (see, e.g., Meyer & Salinger, 1995), even though the definition will only be explicitly used in the proof of Lemma 3.17 in the Appendix. Given a compactly supported wavelet basis $\phi_y = \phi(\cdot - y)$, $y \in \mathbb{Z}$, $\psi_y^j = 2^{j/2}\psi(2^j(\cdot - y))$, $j \ge 0$, $y \in 2^{-j}\mathbb{Z}$, we set

$$
\|g\|_{\mathcal{B}^\beta_{1,\infty}} := \sum_{y\in\mathbb{Z}} |(g, \phi_y)_{L^2}| + \sup_{j\ge 0} 2^{j\beta} \sum_{y\in 2^{-j}\mathbb{Z}} 2^{-j/2}|(g, \psi_y^j)_{L^2}|.
$$

Define $\mathcal{B}^{\beta}_{1,\infty}(\mathbb{R}) := \{g \in L^1 \mid \|g\|_{\mathcal{B}^{\beta}_{1,\infty}} < \infty\}$ for $\beta > 0$, and

$$\mathcal{B}^{\beta}_{1,\infty}(\mathbb{R}) := \left\{ g \in C_c^{-\lceil \beta \rceil + 1}(\mathbb{R}))' \,\middle|\, \|g\|_{\mathcal{B}^{\beta}_{1,\infty}} < \infty \right\}$$

for $\beta \leq 0$.

**Definition 3.5.** Let $\delta^{\varepsilon} : \mathbb{R}^2 \to \mathbb{R}$ be a measurable, bounded function with the following properties:

- $\delta^{\varepsilon}(x, y) = \delta^{\varepsilon}(y, x)$ for all $x, y \in \mathbb{R}$,
- the map $\mathbb{R} \ni x \mapsto \delta^{\varepsilon}(x, \cdot) \in \mathcal{B}^{\beta}_{1,\infty}(\mathbb{R})$ is bounded and measurable for some $\beta > -|\Xi| = 1/2 + \kappa$.
- $\int_{\mathbb{R}} \delta^{\varepsilon}(x, \cdot) \,\mathrm{d}x = 1$,
- $\sup_{x,y \in \mathbb{R}} |\delta^{\varepsilon}(x, y)| \lesssim \varepsilon^{-1}$,
- $\operatorname{supp} \delta^{\varepsilon}(x, \cdot) \subseteq B(x, c \cdot \varepsilon)$ for any $x \in \mathbb{R}$ and some $c > 0$.

**Example 3.6.** There are two examples that are of particular interest for our purposes:

- We say that $\delta^{\varepsilon}$ "comes from a mollifier," if there is symmetric, compactly supported function $\rho \in L^{\infty} \cap \mathcal{B}^{\beta}_{1,\infty}(\mathbb{R})$ integrating to 1 such that

$$\delta^{\varepsilon}(x, y) = \varepsilon^{-1} \cdot \rho(\varepsilon^{-1}(y - x)).$$

- A further interesting example is the case where $\delta^{\varepsilon}$ "comes from a wavelet basis." Consider only $\varepsilon = 2^{-N}$ and choose compactly supported father wavelets $\phi_{k,N} \in L^{\infty} \cap \mathcal{B}^{\beta}_{1,\infty}$, $k \in \mathbb{Z}$ (e.g., the Haar father wavelets $\phi_{k,N} = 2^{N/2} \cdot \mathbf{1}_{[k2^{-N},(k+1)2^{-N})}$) and set

$$\delta^{\varepsilon}(x, y) := \sum_{k \in \mathbb{Z}} \phi_{k,N}(x)\phi_{k,N}(y).$$

  Note that we could also add some generations of mother wavelets in this choice.

Locally, $\dot{W}$ is contained in $\mathcal{B}^{|\Xi|}_{\infty,\infty}(\mathbb{R})$ (recall: $|\Xi| = -1/2 - \kappa$), so that due to $\mathcal{B}^{|\Xi|}_{\infty,\infty}(\mathbb{R}) \subseteq (\mathcal{B}^{\beta}_{1,\infty}(\mathbb{R}))'$, we can define

$$\dot{W}^{\varepsilon}_t := \langle \dot{W}, \delta^{\varepsilon}(t, \cdot) \rangle \mathbf{1}_{\mathbb{R}_+}(t),$$

which is a Gaussian process and pathwise measurable and locally bounded. For (maybe stochastic) integrands $f$, we introduce the notations

$$\int_0^t f(r) \,\mathrm{d}W^{\varepsilon}_r := \int_0^t f(r) \dot{W}^{\varepsilon}_r \,\mathrm{d}r.$$

If $f$ takes values in some (nonhomogeneous) Wiener chaos induced by $\dot{W}$, we also introduce

$$\int_0^t f(r) \diamond \mathrm{d}W^{\varepsilon}_r := \int_0^t f(r) \diamond \dot{W}^{\varepsilon}_r \,\mathrm{d}r, \tag{36}$$

where $\diamond$ denotes the *Wick product*. Note that these two objects do not coincide in general. A complete repetition of the definition of Wick multiplication would stray away too far from the focus of this article. We refer to Janson (1997, Section 3.1) for more details. In essence, a Wick product combines two random variables $X$ and $Y$ (which lie in the a suitable space, namely, the *Wiener chaos*) in a symmetric, bilinear manner to a new random variable $X \diamond Y$ by subtracting from the usual product $X \cdot Y$ a sum of correlation terms. If $X$ and $Y$ are independent, these correlations vanish, so that the Wick product just coincides with the pointwise product $X \cdot Y$. This includes the case of $X$ being constant so that, in particular, $1 \diamond Y = Y$. Another rather simple example arises if $X, Y$ are both (centered) Gaussian random variables. The Wick product is then simply given by

$$X \diamond Y = X \cdot Y - \mathbb{E}[X \cdot Y] .$$

In this article, $X$ and $Y$ are themselves products of Gaussian random variables. In this case, an explicit formula for $X \diamond Y = X \cdot Y - \dots$ is given in equation (3.6) of Janson (1997).

The motivation for using the same symbol "$\diamond$" for Wick products and for Skorokhod integrals (cf. (29)) is that Skorokhod integrals can be seen as "infinitesimal Wick products" in the sense that they are the limit of sums of Wick products, compare Remark 3.7 below. We sketch in Remark 3.7 shortly that one might want to read the Skorokhod integral as

$$\int_0^t f(r) \diamond \mathrm{d}W_r = \}\} \int_0^t f(r) \diamond \frac{\mathrm{d}W}{\mathrm{d}r} \mathrm{d}r," \tag{37}$$

where $\diamond$ should be read as Skorokhod integration on the left and "Wick multiplication" on the right-hand side (which is ill-defined as $\frac{\mathrm{d}W}{\mathrm{d}r}$ only exists as a distribution). The ill-defined identity (37) can be read as a motivation for the (well-defined) definition (36). The close relation between Skorokhod integration and Wick multiplication plays a crucial role in the proof of Theorem 3.14 in the Appendix.

*Remark* 3.7. For the reader's convenience, we briefly comment on the close relation between the Skorokhod integral and the Wick product. Indeed, when $g = \sum X_s \mathbf{1}_{[s,t]}$, with summation over a finite partition of $[0, T]$, and each $X_s$ a (nonadapted) random variable in a finite Wiener–Itô chaos, it follows from (Janson, 1997, Theorem 7.40) that

$$\int g \diamond \mathrm{d}W = \sum X_s \diamond W_{s,t} ,$$

where $\diamond$ denotes Skorokhod integration on the left and Wick multiplication on the right-hand side and where $W_{s,t} = W(t) - W(s)$. Passage to $L^2$-limits is then standard, so that a Skorokhod integral $\int g \diamond \mathrm{d}W$ can be interpreted as the integrated Wick product "$g \diamond \frac{\mathrm{d}W}{\mathrm{d}t}$,"

$$\int g \diamond \mathrm{d}W = \text{“} \int g \diamond \frac{\mathrm{d}W}{\mathrm{d}t} \mathrm{d}t, \text{”}$$

which can be seen as a motivation for our definition (36). See also Nualart (2013) and the references therein.

We now define an approximate fBm by setting

$$\widehat{W}_t^\varepsilon := K \star \dot{W}^\varepsilon(t) = \sqrt{2H} \int_0^t |t - r|^{H-1/2} \, \mathrm{d}W_r^\varepsilon ,$$

which has the expected regularity as it is shown in the following lemma.

**Lemma 3.8.** *On every compact time interval* $[0, T]$*, we have the estimates*

$$|\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon| \lesssim C_\varepsilon |t-s|^{H-\kappa'}, \ |\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon - (\widehat{W}_t - \widehat{W}_s)| \lesssim C|t-s|^{H-\kappa'} \varepsilon^{\delta\kappa'}$$

*uniformly in* $\varepsilon \in (0,1]$ *for any* $\delta \in (0,1)$ *and* $\kappa' \in (0, H)$*. Here,* $C_\varepsilon, C > 0$ *are random constants that are (uniformly) bounded in* $L^p$ *for* $p \in [1, \infty)$*.*

*Proof.* The proof is elementary but a bit bulky and therefore postponed to the Appendix. □

Finally, we can give the definition of the approximating model $(\Pi^\varepsilon, \Gamma^\varepsilon)$, the "canonical" model built from the approximate (and hence regular) noise $W^\varepsilon$.

$$\begin{array}{ll}
\Pi_s^\varepsilon \mathbf{1} = 1, & \Gamma_{st}^\varepsilon \mathbf{1} = 1, \\
\Pi_s^\varepsilon \Xi = \dot{W}^\varepsilon, & \Gamma_{st}^\varepsilon \Xi = \Xi, \\
\Pi_s^\varepsilon \mathcal{I}(\Xi)^m = \left(\widehat{W}_\cdot^\varepsilon - \widehat{W}_s^\varepsilon\right)^m, & \Gamma_{st}^\varepsilon \mathcal{I}(\Xi) = \mathcal{I}(\Xi) + \left(\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon\right)\mathbf{1}, \\
\Pi_s^\varepsilon \mathcal{I}(\Xi)^m \Xi = (\widehat{W}_\cdot^\varepsilon - \widehat{W}_s^\varepsilon)^m \, \dot{W}_\cdot^\varepsilon, & \Gamma_{st}^\varepsilon \tau\tau' = \Gamma_{st}^\varepsilon \tau \cdot \Gamma_{st}^\varepsilon \tau', \quad \text{for } \tau, \tau' \in S \text{ with } \tau\tau' \in S.
\end{array}$$

**Lemma 3.9.** *The pair* $(\Pi^\varepsilon, \Gamma^\varepsilon)$ *as defined above is a model on* $(\mathcal{T}, A)$*.*

*Proof.* The identity $\Pi_t = \Gamma_{ts}\Pi_s$ is straightforward to check. The bounds (34) and (35) on $\Gamma_{st}$ and on $\Pi_s \mathcal{I}(\Xi)^m$ follow from the regularity of $\widehat{W}^\varepsilon$ as proved in Lemma 3.8. The blowup of $\Pi_s \Xi \mathcal{I}(\Xi)^m(\varphi_s^\lambda)$, however, is even better than we need, because by the choice of $\delta^\varepsilon$, we have $|\dot{W}^\varepsilon| \leq C_\varepsilon$ on compact sets, for some random constant $C_\varepsilon$. □

The definition of this model is justified by the fact that application of the reconstruction operator (as in Lemma 3.23) yields integrals of the form

$$\int_0^t f(r, \widehat{W}_r^\varepsilon) \, \mathrm{d}W_r^\varepsilon \,. \tag{38}$$

As we pointed out already in Section 1, there is no hope that integrals of this type will converge as $\varepsilon \to 0$ if $H < 1/2$. This can be cured by working with a renormalized model $(\widehat{\Pi}^\varepsilon, \Gamma^\varepsilon)$ instead.

### 3.1.3 | The renormalized model $\widehat{\Pi}^\varepsilon$

From the perspective of regularity structures, the fundamental reason why integrals like (38) fail to converge to

$$\int_0^t f(r, \widehat{W}_r) \, \mathrm{d}W_r$$

lies in the fact that the corresponding models will not satisfy $(\Pi^\varepsilon, \Gamma^\varepsilon) \to (\Pi, \Gamma)$ in a suitable norm. To see what is going on, we will first rewrite $\Pi_s \Xi \mathcal{I}(\Xi)^k$.

**Lemma 3.10.** *For* $\varphi \in C_c^\infty(\mathbb{R})$*,* $s \in \mathbb{R}$*,* $m \in \{1, \dots, M\}$*, we have*

$$\begin{aligned}
\Pi_s \Xi \mathcal{I}(\Xi)^m(\varphi) = &\int_0^\infty \varphi(t) \, (\widehat{W}_t - \widehat{W}_s)^m \diamond \mathrm{d}W_t \\
&- m \int_0^\infty \varphi(t) \, K(s-t) \, (\widehat{W}_t - \widehat{W}_s)^{m-1} \, \mathrm{d}t,
\end{aligned}$$

*where $\diamond$ denotes the Skorokhod integral and $K(t) = \sqrt{2H}\mathbf{1}_{t>0}t^{H-1/2}$ denotes the Volterra kernel. Note that in the second term, the domain of integration is actually* $(0, s)$.

*Remark* 3.11. At first glance, it might seem surprising (and maybe confusing) that we are in need of a Skorokhod integral to write down the integrals in Lemma 3.10, especially for a reader more familiar with rough path theory. Note that the expression $\int_0^\infty \varphi(t)\,(\widehat{W}_t - \widehat{W}_s)^m \diamond \mathrm{d}W_t$ is *ill-defined as an Itô integral*: As we allow $\varphi$ to have support below $s$, the domain of integration involves $t$ with $t < s$, in which case $(\widehat{W}(t) - \widehat{W}(s))^m$ is not adapted with respect to the filtration of $W$ at time $t$. The concept of a Skorokhod integral is in such cases a natural extension of Itô's notion of integration that boils down to the classical Itô integral once the integrand is adapted. In the theory of rough paths, one usually takes $\varphi(t) = \mathbf{1}_{[s,t']}(t)$ for $s < t'$ (see also P. K. Friz & Hairer, 2014, Section 13.3.2), which explains why issues of this kind never arise in the rough path framework.

There is, in fact, a way to write the integral $\int_0^\infty \varphi(t)\,(\widehat{W}_t - \widehat{W}_s)^m \diamond \mathrm{d}W_t$ as an Itô integral: Expand $(\widehat{W}_t - \widehat{W}_s)^m$ via the binomial theorem and "pull out" all factors depending on $s$ (such a point of view is essentially behind Chen's relation in (32)). However, as this seems like a rather nebulous definition, we consider it more convenient to work with a more appropriate notion of integration in this article.

*Proof of Lemma* 3.10. We prove this by reexpressing $\mathbb{W}^k_{s,t}$. For $s < t$, we already have

$$\mathbb{W}^k_{s,t} = \int_s^t \mathrm{d}W_r \diamond (\widehat{W}_r - \widehat{W}_s)^k,$$

so that it remains to see what happens for $t < s$. With relation (32), we then have

$$\mathbb{W}^k_{s,t} = -\sum_{l=0}^{k} \binom{k}{l} (\widehat{W}_t - \widehat{W}_s)^l \cdot \int_{\mathbb{R}} \mathrm{d}r\,\dot{W}_r \diamond (\widehat{W}_r - \widehat{W}_t)^{k-l}\mathbf{1}_{t<r<s}\,,$$

where we use for the sake of concision formal notation, which is easy to translate to a rigorous formulation. Using the fact that for Gaussian random variables $U_1, V, U_2$, we have

$$U_1^l \cdot (V \diamond U_2^{k-l}) = V \diamond (U_1^l U_2^{k-l}) + l\mathbb{E}[VU_1]\,U_1^{l-1}U_2^{k-l} \tag{39}$$

(a consequence of Janson, 1997, Theorems 3.15 and 7.33), we obtain

$$\begin{aligned}\mathbb{W}^k s, t = &-\int_{\mathbb{R}} \mathrm{d}r\,\dot{W}_r \diamond (\widehat{W}_r - \widehat{W}_s)^k \mathbf{1}_{t<r<s} \\ &- \sum_{l=0}^{k} \binom{k}{l} l \cdot \int_{\mathbb{R}} \mathrm{d}r\,\mathbb{E}[\dot{W}_r \cdot (\widehat{W}_t - \widehat{W}_s)] \cdot (\widehat{W}_t - \widehat{W}_s)^{l-1} \cdot (\widehat{W}_r - \widehat{W}_t)^{k-l}\,.\end{aligned}$$

Using $\binom{k}{l} = k\binom{k-1}{l-1}$ and $\mathbb{E}[\dot{W}(r) \cdot (\widehat{W}(t) - \widehat{W}(s))] = -K(s-r)\mathbf{1}_{r>0}$ for $t < r < s$, we can reformulate this expression and obtain

$$\mathbb{W}^k(s,t) = -\int \mathrm{d}W(r) \diamond (\widehat{W}(r) - \widehat{W}(s))^k \mathbf{1}_{t<r<s} + k\int \mathrm{d}rK(s-r)(\widehat{W}(r) - \widehat{W}(s))^{k-1}\mathbf{1}_{r>0}\,.$$

(An alternative derivation can be given following Nualart & Pardoux, 1988, Theorem 3.2.) As $\Pi_s \Xi \mathcal{I}(\Xi)^m(\varphi) = \int_{\mathbb{R}} \varphi(t)\,\mathrm{d}_t \mathbb{W}^m_{s,t}$, the claim follows. □

Let us also reexpress the approximating model in suitable form.

**Lemma 3.12.** *For $\varphi \in C_c^\infty(\mathbb{R})$, $s \in \mathbb{R}$, $m \in \{1, \dots, M\}$, we have*

$$\begin{aligned}\Pi_s^\varepsilon \Xi \mathcal{I}(\Xi)^m(\varphi) = &\int_0^\infty \varphi(t)\,(\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^m \diamond \mathrm{d}W_t^\varepsilon \\ &- m\int_0^\infty \varphi(t)\,\mathscr{K}^\varepsilon(s,t)(\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^{m-1}\mathrm{d}t \\ &+ m\int_0^\infty \varphi(t)\,\mathscr{K}^\varepsilon(t,t)(\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^{m-1}\,\mathrm{d}t,\end{aligned}$$

*where $\diamond$ is defined as in (36) and where*

$$\mathscr{K}^\varepsilon(u,v) := \mathbb{E}[\widehat{W}^\varepsilon(u)\dot{W}^\varepsilon(v)] = \mathbf{1}_{u,v\ge 0}\int_0^\infty\int_0^\infty \delta^\varepsilon(v,x_1)\delta^\varepsilon(x_1,x_2)K(u-x_2)\,\mathrm{d}x_1\mathrm{d}x_2\,. \tag{40}$$

*Proof.* Using that for Gaussian random variables $V$, $U$ we have $VU^m = V \diamond U^m + m\mathbb{E}[VU]U^{m-1}$ (see (39) with $U_2 = 1$), we can rewrite

$$\begin{aligned}\Pi_s^\varepsilon \Xi \mathcal{I}(\Xi)^m(\varphi) = &\int_0^\infty \varphi(t)\,(\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^m \diamond \mathrm{d}W_t^\varepsilon \\ &+ m\int_0^\infty \mathrm{d}t\,\varphi(t)\,\mathbb{E}[\dot{W}_t^\varepsilon\,(\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)](\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^{m-1}.\end{aligned}$$

Inserting $\mathbb{E}[\dot{W}^\varepsilon(t)\,(\widehat{W}^\varepsilon(t) - \widehat{W}^\varepsilon(s))] = \mathscr{K}^\varepsilon(t,t) - \mathscr{K}^\varepsilon(s,t)$ shows the identity. □

Comparing the expressions in Lemmas 3.12 and 3.10, we see that we should subtract

$$m\int \varphi(t)\,\mathscr{K}^\varepsilon(t,t)(\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^{m-1}\,\mathrm{d}t$$

from the model, which will give us a new model $\widehat{\Pi}^\varepsilon$. Of course, we have to be careful that this step preserves "Chen's relation" $\widehat{\Pi}_s^\varepsilon\Gamma_{st} = \widehat{\Pi}_t^\varepsilon$, see Theorem 3.14 below.

If we interpret $\mathscr{K}^\varepsilon$ as an approximation to the Volterra kernel, we see that the expression

$$\mathscr{C}^\varepsilon(t) := \mathscr{K}^\varepsilon(t,t),\; t \ge 0,$$

will correspond to something like "$0^{H-1/2} = \infty$" in the limit $\varepsilon \to 0$. We indeed have the following upper bound.

**Lemma 3.13.** *For all $s, t \in \mathbb{R}$, we have*

$$|\mathscr{K}^\varepsilon(s,t)| \lesssim \varepsilon^{H-1/2}\,.$$

*Proof.* $|\mathscr{K}^\varepsilon(s,t)| \lesssim \varepsilon^{-2}\int_{B(t,c\varepsilon)}\mathrm{d}x\int_{B(x,c\varepsilon)}\mathrm{d}u\,|s-u|^{H-1/2} \lesssim \varepsilon^{H-1/2}\,.$ □

Our hope is now that the new model $\widehat{\Pi}^\varepsilon$ converges to $\Pi$ in a suitable sense. Similar to Hairer (2014, (2.17)), we define the distance between two models $(\Pi,\Gamma)$ and $(\widetilde{\Pi},\widetilde{\Gamma})$ on a compact time interval

$[0, T]$ as

$$|||(\Pi, \Gamma); (\widetilde{\Pi}, \widetilde{\Gamma})|||_T := \sup_{\substack{\operatorname{supp}\varphi \subseteq B(0,1),\\ \lambda \in (0,1],\\ s \in [0,T], \tau \in S}} \lambda^{-|\tau|} |(\Pi_s - \widetilde{\Pi}_s)\tau(\varphi_s^\lambda)| + \sup_{\substack{t,s \in [0,T],\\ \tau \in S, A \ni \beta < |\tau|}} \frac{|\Gamma_{ts}\tau - \widetilde{\Gamma}_{ts}\tau|_\beta}{|t-s|^{|\tau|-\beta}}, \tag{41}$$

where $|\cdot|_\beta$ denotes the absolute value of the coefficient of the symbol $\tau'$ with $|\tau'| = \beta$ and where the first supremum runs over $\varphi \in C_c^1$ with $\|\varphi\|_{C^1} \leq 1$. We will also need

$$\|\Pi\|_T = \sup_{\substack{\operatorname{supp}\varphi \subseteq B(0,1),\\ \lambda \in (0,1],\\ s \in [0,T], \tau \in S}} \lambda^{-|\tau|} |\Pi_s \tau(\varphi_s^\lambda)| .$$

We are now ready to give the fundamental result of this subsection. Recall that the (minimal) homogeneity is $|\Xi| = -1/2 - \kappa$, which corresponds to $W$ being Hölder with exponent $1/2 - \kappa$.

**Theorem 3.14.** *For every $s \in [0, T]$, define the linear map $\widehat{\Pi}_s^\varepsilon : \mathcal{T} \to C_c^1(\mathbb{R})'$ given by*

$$\widehat{\Pi}_s^\varepsilon \Xi \mathcal{I}(\Xi)^m = \Pi_s^\varepsilon \Xi \mathcal{I}(\Xi)^m - m\mathscr{C}^\varepsilon(\cdot)\Pi_s^\varepsilon(\mathcal{I}(\Xi)^{m-1}), \quad m \in \{1, \dots, M\},$$

*and $\widehat{\Pi}_s^\varepsilon = \Pi_s^\varepsilon$ on all remaining symbols in $S$. Then,*

$$(\widehat{\Pi}^\varepsilon, \widehat{\Gamma}^\varepsilon) := (\widehat{\Pi}^\varepsilon, \Gamma^\varepsilon)$$

*defines a ("renormalized") model on $(\mathcal{T}, A)$. On compact time intervals, we have*

$$\Big\| |||(\widehat{\Pi}^\varepsilon, \widehat{\Gamma}^\varepsilon); (\Pi, \Gamma)|||_T \Big\|_{L^p} \lesssim \varepsilon^{\delta\kappa}, \tag{42}$$

*for any $\delta \in (0,1)$ and $p \in [1, \infty)$. In particular, the distance between the renormalized model and the Itô model almost decays with rate $H$ for $M = M(\kappa, H)$ large enough.*

*Remark* 3.15. In the special case of the level-2 Brownian rough path (i.e., $H = 1/2$, $M = 1$ ), the above result is in precise agreement with known results—but note that we are dealing with the simple case of scalar Brownian motion. More specifically, we do not see the usual (strong) rate "almost" $1/2$, but have to subtract the Hölder exponent used in the rough path/model topology (here: $1/2 - \kappa$), which almost leads to the rate $\kappa$. As $M = 1$ entails the condition $1/2 - \kappa > 1/3$, we see that $\kappa < 1/6$, exactly as given, for example, in P. K. Friz and Hairer (2014, Ex. 10.14). A better rate can be achieved by working with higher level rough paths (here: $M > 1$), and indeed, the special case of $H = 1/2$, but general $M$, can be seen as a consequence of P. Friz and Riedel (2011): at the price of working with $\sim 1/(1/2 - \kappa)$ levels, one can choose $\kappa$ arbitrarily close to $1/2$ and so recover the usual "almost" $1/2$ rate. Of course, the case $H < 1/2$ is out of reach of rough path considerations.

*Proof.* Due to Lemma 3.13 we have, for fixed $\varepsilon$, that $\sup_{t\in[0,T]} |\mathscr{C}^\varepsilon(t)| < \infty$. As $|\Pi_s \mathcal{I}(\Xi)^m| \lesssim |\cdot - s|^{mH}$, the bound (34) is still satisfied. The modification $\widehat{\Pi}_s^\varepsilon \Xi \mathcal{I}(\Xi)^m - \Pi_s^\varepsilon \Xi \mathcal{I}(\Xi)^m$ does not lead to a violation of "Chen's relation." Indeed, using validity of (33) for the original model, we have

$$\hat{\Pi}_t^\varepsilon \Gamma_{ts}^\varepsilon(\Xi \mathcal{I}(\Xi)^k) = \hat{\Pi}_t^\varepsilon \left( \sum_{l=0}^{k} \binom{k}{l} (\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^l \Xi \mathcal{I}(\Xi)^{k-l} \right)$$

$$= \Pi_s^\varepsilon(\Xi \mathcal{I}(\Xi)^k) - \sum_{l=0}^{k} \binom{k}{l} (\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^l (k-l) \mathscr{C}^\varepsilon(\cdot)(\widehat{W}_\cdot^\varepsilon - \widehat{W}_t^\varepsilon)^{k-l-1}$$

$$= \Pi_s^\varepsilon(\Xi \mathcal{I}(\Xi)^k) - k\mathscr{C}^\varepsilon(\cdot) \sum_{l=0}^{k-1} \binom{k-1}{l} (\widehat{W}_t^\varepsilon - \widehat{W}_s^\varepsilon)^l \, (\widehat{W}_\cdot^\varepsilon - \widehat{W}_t^\varepsilon)^{k-l-1}$$

$$= \Pi_s^\varepsilon(\Xi \mathcal{I}(\Xi)^k) - k\mathscr{C}^\varepsilon(\cdot)(\widehat{W}_\cdot^\varepsilon - \widehat{W}_s^\varepsilon)^k = \hat{\Pi}_s^\varepsilon(\Xi(\mathcal{I}(\Xi)^k)\,.$$

We see that (33) is also satisfied after our modification, and then easily conclude that $(\hat{\Pi}^\varepsilon, \Gamma^\varepsilon)$ is still a model on $(\mathcal{T}, A)$. At last, the bound (42) is a bit technical and left to Appendix A. □

## 3.2 | Approximation and renormalization theory

We now address the central question of how the integral $\int_0^t f(\widehat{W}_r^\varepsilon, r)\,\mathrm{d}W_r^\varepsilon$ has to be modified to make it convergent to $\int_0^t f(W_r, r)\mathrm{d}W_r$. The key idea is to combine the convergence result from Theorem 3.14 with Hairer's reconstruction theorem, which we state below.

We first recall the notion of a modeled distribution, compare Hairer (2014, Definition 3.1). We say that a map $F : \mathbb{R} \to \mathcal{T}$ is in the space $\mathcal{D}_T^\gamma(\Gamma)$, $\gamma > 0$, for some time horizon $T > 0$ if

$$\|F\|_{\mathcal{D}_T^\gamma(\Gamma)} := \sup_{A \ni \beta < \gamma, s \in [0,T]} |F(s)|_\beta + \sup_{A \ni \beta < \gamma, s,t \in [0,T],\, s \neq t} \frac{|F(t) - \Gamma_{ts} F(s)|_\beta}{|t-s|^{\gamma-\beta}} < \infty, \tag{43}$$

where, as above, $|\cdot|_\beta$ denotes the absolute value of the coefficient of $\tau$ with $|\tau| = \beta$. Given two models $(\Pi, \Gamma)$ and $(\overline{\Pi}, \overline{\Gamma})$ and two $F, \overline{F} : \mathbb{R} \mapsto \mathcal{T}$, it is also useful to have the notion of a distance

$$|\!|\!|F; \overline{F}|\!|\!|_{\mathcal{D}_T^\gamma(\Gamma), \mathcal{D}_T^\gamma(\overline{\Gamma})} := \sup_{A \ni \beta < \gamma,\, t \in [0,T]} |F(t) - \overline{F}(t)|_\beta$$

$$+ \sup_{A \ni \beta < \gamma,\, s,t \in [0,T],\, s \neq t} \frac{|F(t) - \Gamma_{ts}F(s) - (\overline{F}(t) - \overline{\Gamma}_{ts}\overline{F}(s))|_\beta}{|t-s|^{\gamma-\beta}}\,.$$

The reconstruction theorem now states that for $\gamma > 0$, a map $F \in \mathcal{D}_T^\gamma(\Gamma)$ can be uniquely identified with a distribution that behaves locally like $\Pi_\cdot F(\cdot)$.

**Theorem 3.16.** *(Hairer, 2014, Theorem 3.10). Given a model* $(\Pi, \Gamma)$*,* $\gamma > 0$*, and* $T > 0$*, there is a unique continuous operator*[4] $\mathcal{R} : \mathcal{D}_T^\gamma(\Gamma) \to C^{|\Xi|}(\mathbb{R})$ *such that for any* $s \in [0,T]$ *and* $\varphi \in C_c^1(B(0,1))$

$$|(\mathcal{R}F - \Pi_s F(s))(\varphi_s^\lambda)| \lesssim \|\Pi\|_T\, \lambda^\gamma\,. \tag{44}$$

*For two different models* $(\Pi, \Gamma)$ *and* $(\overline{\Pi}, \overline{\Gamma})$*, we further have*

$$\left|(\mathcal{R}F - \Pi_s F(s) - (\overline{\mathcal{R}}\,\overline{F} - \Pi_s \overline{F}(s)))(\varphi_s^\lambda)\right|$$

$$\lesssim \lambda^\gamma \left( \|F\|_{\mathcal{D}_T^\gamma(\Gamma)} |\!|\!|(\Pi,\Gamma); (\overline{\Pi}, \overline{\Gamma})|\!|\!|_T + \|\Pi\|_T |\!|\!|F; \overline{F}|\!|\!|_{\mathcal{D}_T^\gamma(\Gamma); \mathcal{D}_T^\gamma(\overline{\Gamma})} \right) \tag{45}$$

*for* $F \in \mathcal{D}_T^\gamma(\Gamma)$*,* $\overline{F} \in \mathcal{D}_T^\gamma(\overline{\Gamma})$*.*

As mentioned earlier, we prefer to work with compactly supported functions $\varphi \in \mathcal{B}^{\beta}_{1,\infty}(\mathbb{R}^d)$, $\beta > -|\Xi|$, which includes objects like the Haar wavelets. The following lemma allows us to carry over all bounds.

**Lemma 3.17.** *The bounds (34), (41), (44), and (45) still hold for $\varphi \in \mathcal{B}^{\beta}_{1,\infty}(\mathbb{R}^d)$, $\beta > -|\Xi|$, with compact support in $B(0,1)$ (after a change of constants).*

*Remark* 3.18. This covers in particular functions like $\mathbf{1}_{[0,1]} \in \mathcal{B}^{1}_{1,\infty}(\mathbb{R})$.

*Proof.* We prove this via wavelet methods in the Appendix. □

In the following, we introduce $X^{(\varepsilon)}$ to denote both $X$ and $X^{\varepsilon}$. To study objects like $\int_0^t f(\widehat{W}^{(\varepsilon)}_r, r)\,\mathrm{d}W^{(\varepsilon)}_r$ with the reconstruction theorem, we first "expand" the integrand $f(\widehat{W}^{(\varepsilon)}_r, r)$ in the regularity structure $\mathcal{T}$, obtaining

$$F^{(\varepsilon)}(s) := \sum_{m=0}^{M} \frac{1}{m!} \partial_1^m f(\widehat{W}^{(\varepsilon)}_s, s)\mathcal{I}(\Xi)^m.$$

On the level of the regularity structure, these objects can be multiplied with the "noise" $\Xi$, which gives a modeled distribution on $\mathcal{T}$. We will analyze $F^{(\varepsilon)}$ by writing it as the composition of a (random) modeled distribution with the smooth function $f$. To this end, we need the following.

**Lemma 3.19.** *On the regularity structure $(\mathcal{T}, A, G)$ introduced in Section 3.1, consider a model $(\Pi, \Gamma)$ that is* admissible *in the sense that*

$$\Pi_t \mathcal{I}(\Xi) = (K * \Pi_t \Xi)(\cdot) - (K * \Pi_t \Xi)(t)\,.$$

*Then*

$$\mathcal{K}\Xi(t) := \mathcal{I}(\Xi) + (K * \Pi_t \Xi)(t)\mathbf{1} \tag{46}$$

*defines a modeled distribution. More precisely, $\mathcal{K}\Xi \in \mathcal{D}^{\infty}_T := \bigcup_{\gamma<\infty} \mathcal{D}^{\gamma}_T$.*

*Remark* 3.20. Our notion of admissibility mimics (Hairer, 2014, Definition 5.9), which, however, is not directly applicable here (due to the failure of assumption 5.4 in Hairer, 2014).

*Proof.* By definition of the space of modeled distribution, we need to understand the action of $\Gamma_{st}$ on all constituting symbols. As $\{\mathbf{1}, \mathcal{I}(\Xi)\}$ span a sector, that is, a space invariant by the action of the structure group, it is clear that

$$\Gamma_{st}\mathcal{I}(\Xi) = \mathcal{I}(\Xi) + (\ldots)\mathbf{1}.$$

Application of the realization map $\Pi_s$, followed by evaluation at $s$, immediately identifies $(\ldots.)$ with

$$\Pi_t \mathcal{I}(\Xi)(s) - \Pi_s \mathcal{I}(\Xi)(s) = \Pi_t \mathcal{I}(\Xi)(s) = (K * \Pi_s \Xi)(s) - (K * \Pi_t \Xi)(t),$$

where we used admissibility and $\Pi_s \Xi = \Pi_t \Xi$ in the last step. As a consequence, $\Gamma_{st}\mathcal{K}\Xi(t) \equiv \mathcal{K}\Xi(s)$ so that trivially $\mathcal{K}\Xi \in \mathcal{D}^{\gamma}_T$ for any $\gamma < \infty$. □

For a given (sufficiently smooth) function $f$, and a generic model $(\Pi, \Gamma)$ on our regularity structure, define

$$F^{\Pi} : s \mapsto \sum_{m=0}^{M} \frac{1}{m!} \partial_1^m f((\mathcal{R}\mathcal{K}\Xi(s), s)\mathcal{I}(\Xi)^m.$$

Remark that $\mathcal{K}\Xi(s)$ is function-like, that is, takes values in the span of symbols with nonnegative degree. From Hairer (2014, Proposition 3.28), we then have

$$\mathcal{R}\mathcal{K}\Xi(s) = \langle \mathcal{K}\Xi(s), \mathbf{1} \rangle = K * \Pi_s \Xi.$$

(In particular, we see that $F^{(\varepsilon)}(s)$ coincides with $F^{\Pi}$ when $\Pi$ is taken as either the approximate or the renormalized approximate model.) We can also define $\Xi F^{\Pi}$ simply obtained by multiplying $F^{\Pi}$ with $\Xi$. The properties of $F^{\Pi}$ and $\Xi F^{\Pi}$ are summarized in the following lemma.

**Lemma 3.21.** *Given $f \in C_b^{2M+3}([0,T] \times \mathbb{R})$, there exists $N > 0$ such that, for all $\gamma \in (1/2 + \kappa, 1)$,*

$$\|F^{\Pi}\|_{\mathcal{D}_T^{\gamma}(\Gamma)} \lesssim \|\Pi\|_T^N, \quad \|\Xi F^{\Pi}\|_{\mathcal{D}_T^{\gamma+|\Xi|}(\Gamma)} \lesssim \|\Pi\|_T^N.$$

*We further have, for two given models $(\Pi, \Gamma)$ and $(\Pi', \Gamma')$,*

$$|\!|\!| F^{\Pi}; F^{\Pi'} |\!|\!|_{\mathcal{D}_T^{\gamma}(\Gamma);\mathcal{D}_T^{\gamma}(\Gamma')} \lesssim \left( \|\Pi\|_T^N + \|\Pi'\|_T^N \right) |\!|\!| (\Pi,\Gamma); (\Pi',\Gamma') |\!|\!|_T, \tag{47}$$

$$|\!|\!| \Xi F^{\Pi}; \Xi F^{\Pi'} |\!|\!|_{\mathcal{D}_T^{\gamma+|\Xi|}(\Gamma);\mathcal{D}_T^{\gamma+|\Xi|}(\Gamma')} \lesssim \left( \|\Pi\|_T^N + \|\Pi'\|_T^N \right) |\!|\!| (\Pi,\Gamma); (\Pi',\Gamma') |\!|\!|_T, \tag{48}$$

*where the proportionality constants are, in particular, uniform over all $f$ with bounded $C^{2M+3}$-norm.*

*Proof.* The map $F^{\Pi}$ is simply the composition (in the sense of Hairer, 2014, Section 4.2) of the function $f$ with the modeled distributions $\mathcal{K}\Xi$ and $s \mapsto s\mathbf{1}$. The result then follows from Hairer (2014, Theorem 4.16) (polynomial dependence in $\|\Pi\|_T$ is not stated there but is clear from the proof). □

*Remark* 3.22. In the case when $f \in C^{2M+3}$, but with no global bounds, the result still holds as we only consider the values of $f$ on the range of the continuous function $\mathcal{R}\mathcal{K}\Xi$ (which is bounded by some $R \geq 0$). The resulting bounds then depend linearly on $\|f\|_{C^{2M+3}(B_R \times [0,T])}$.

In the case of the Itô model $(\Pi, \Gamma)$ and the approximating renormalized models $(\widehat{\Pi}^{\varepsilon}, \Gamma^{\varepsilon})$, we simply denote $F^{\Pi}$ by $F$ and $F^{\varepsilon}$, respectively. We are then allowed to apply Hairer's reconstruction theorem (see Theorem 3.16). Note that we have two reconstruction operators $\mathcal{R}$ and $\mathcal{R}^{\varepsilon}$, because we start with two models. $\mathcal{R}^{(\varepsilon)}\Xi F^{(\varepsilon)}$ can be written down explicitly.

**Lemma 3.23.** *We have (a.s.)*

$$\mathcal{R}F\Xi(\varphi) = \int_{\mathbb{R}} \varphi(t)\, f(\widehat{W}_t, t)\, \mathrm{d}W_t\,,$$

$$\mathcal{R}^{\varepsilon} F^{\varepsilon}\Xi(\varphi) = \int_{\mathbb{R}} \varphi(t)\, f(\widehat{W}_t^{\varepsilon}, t)\, \mathrm{d}W_t^{\varepsilon} - \int_{\mathbb{R}} \mathscr{K}^{\varepsilon}(t,t)\partial_1 f(\widehat{W}_t^{\varepsilon}, t)\varphi(t)\, \mathrm{d}t\,.$$

*Proof.* The proof is in the Appendix. □

If we take $\varphi = \mathbf{1}_{[0,T)}$, we obtain $\mathcal{R}F\Xi(\mathbf{1}_{[0,T)}) = \int_0^T f(\widehat{W}(t), t)\, \mathrm{d}W(t)$, so that it is natural to choose $\tilde{\mathscr{I}}_f^{\varepsilon}(T) = \mathcal{R}^{\varepsilon}\Xi F^{\varepsilon}(\mathbf{1}_{[0,T)})$ as an approximation. However, note that the key property of the reconstruction operator $\mathcal{R}^{(\varepsilon)}$ is that it is locally close to the corresponding model $\Pi^{(\varepsilon)}$, so that we, in fact, have two natural approximations:

**Definition 3.24.** For $F^\varepsilon$ as in Lemma 3.21 and $t \geq 0$, we set

$$\widetilde{\mathscr{I}}_f^\varepsilon(t) := \mathcal{R}^\varepsilon \Xi F^\varepsilon(\mathbf{1}_{[0,t]}) = \int_0^t f(\widehat{W}_r^\varepsilon, r)\,\mathrm{d}W_r^\varepsilon - \int_0^t \mathscr{C}^\varepsilon(r)\partial_1 f(\widehat{W}_r^\varepsilon, r)\,\mathrm{d}r\,.$$

For a fixed partitions $\{[t_l^\varepsilon, t_{l+1}^\varepsilon)\}$ of $[0,t)$ with $|t_{l+1}^\varepsilon - t_l^\varepsilon| \lesssim \varepsilon$, we further set

$$\begin{aligned}\widetilde{\mathscr{J}}_{f,M}^\varepsilon(t) &:= \sum_{[t_l^\varepsilon, t_{l+1}^\varepsilon)} \widehat{\Pi}_{t_l}^\varepsilon \Xi F_{t_l}^\varepsilon(\mathbf{1}_{[t_l^\varepsilon, t_{l+1}^\varepsilon)}) \\ &= \sum_{[t_l^\varepsilon, t_{l+1}^\varepsilon)} \sum_{m=0}^M \frac{1}{m!}\partial_1^m f(\widehat{W}^\varepsilon(t_l^\varepsilon), t_l^\varepsilon) \int_{t_l^\varepsilon}^{t_{l+1}^\varepsilon} \left(\widehat{W}^\varepsilon(r) - \widehat{W}^\varepsilon(t_l^\varepsilon)\right)^m \mathrm{d}W^\varepsilon(r) \\ &\quad - \sum_{m=1}^M \frac{1}{(m-1)!}\partial_1^m f(\widehat{W}^\varepsilon(t_l^\varepsilon), t_l^\varepsilon) \int_{t_l^\varepsilon}^{t_{l+1}^\varepsilon} \mathscr{C}^\varepsilon(r) \left(\widehat{W}^\varepsilon(r) - \widehat{W}^\varepsilon(t_l^\varepsilon)\right)^{m-1} \mathrm{d}r.\end{aligned}$$

We might drop the indices $f$ and $f, M$ on $\widetilde{\mathscr{I}}^\varepsilon$ and $\widetilde{\mathscr{J}}^\varepsilon$ if there is no risk of confusion.

The following theorem, which can be seen as the fundamental theorem of our regularity structure approach to rough pricing, shows that these approximations do both converge.

**Theorem 3.25.** *Fix $T > 0$. For $f$ smooth, bounded with bounded derivatives, and $\widetilde{\mathscr{I}}_f^\varepsilon$, $\widetilde{\mathscr{J}}_{f,M}^\varepsilon$ as in Definition 3.24, we have*

*(i) for any $\delta \in (0,1)$ and any $p < \infty$, there exists $C$ such that*

$$\left\| \sup_{t\in[0,T]} \left| \widetilde{\mathscr{I}}_f^\varepsilon(t) - \int_0^t f(\widehat{W}_r, r)\mathrm{d}W_r \right| \right\|_{L^p} \leq C\varepsilon^{\delta H}\,, \tag{49}$$

*(ii) for every $\delta \in (0,1)$, we can pick $M = M(\delta, H)$ large enough, such that for any $p < \infty$, there exists $C$ such that*

$$\left\| \sup_{t\in[0,T]} \left| \widetilde{\mathscr{J}}_{f,M}^\varepsilon(t) - \int_0^t f(\widehat{W}_r, r)\mathrm{d}W_r \right| \right\|_{L^p} \leq C\varepsilon^{\delta H}\,. \tag{50}$$

*Remark* 3.26. With regard to (i), although $\widetilde{\mathscr{I}}_f^\varepsilon(t)$ does not depend on the choice of $M$, and nor does its (Itô) limit, the choice of $M$ affects the entire regularity structure and so, implicitly also the reconstruction operator $\mathcal{R}^\varepsilon$ used in the definition of $\widetilde{\mathscr{I}}_f^\varepsilon$, as well as the modeled distribution $F^\varepsilon$. The latter, in turn, requires $f \in C^M$ for the construction to make sense. If $\delta$ is chosen arbitrarily close to 1, $f$ needs to have derivatives of arbitrary order, hence our smoothness assumption.

*Remark* 3.27. By an easy localization argument, one shows that for $f$ smooth (but without any further bounds), one still has

$$\sup_{\varepsilon\in(0,1]} \mathbb{P}\left( \sup_{t\in[0,T]} \left| \widetilde{\mathscr{I}}_f^\varepsilon(t) - \int_0^t f(\widehat{W}_r, r)\mathrm{d}W_r \right| \leq C\varepsilon^{\delta H} \right) \to 0$$

with $C \to \infty$. The original rough volatility model due to Gatheral et al. (2018) makes a point that $f$ should be of exponential form. Now, the result with $L^p$-estimates still holds because we only consider the values of $f$ on the range of the continuous function $\mathcal{R}\mathcal{K}\Xi$ (which is bounded by some $R \geq 0$). As pointed out in Remark 3.22, the bounds then depend linearly on $\|f\|_{C^{M+2}(B_R\times[0,T])}$. Behause $(\Pi, \Gamma)$ is Gaussian model, $\mathcal{R}\mathcal{K}\Xi$ is a Gaussian process (say, $\widehat{W}$ or $\widehat{W}^\varepsilon$), and hence we have Gaussian concentration of Fernique-type for $\sup_{t\in[0,T]} |\mathcal{R}\mathcal{K}\Xi(t)|$. So, for instance, if $f$ and its derivatives have exponential growth, we do have the $L^p$ bounds of the above theorem, for all $p < \infty$. This remark justifies in particular the choice $f(x) = \exp(x)$ and $p = 2$ in the numerical discussion of Section 6.

*Remark* 3.28. In Neuenkirch and Shalaiko (2016), it is shown (in a slightly different setting) that the strong rate for the standard Euler scheme (or, more precisely, left-point rule) is no better than $H$ in general even when the fractional process is exactly simulated. In that sense, the scheme suggested in Theorem 3.25 is almost optimal.

*Proof of Theorem* 3.25. Without loss of generality $T \leq 1$, otherwise split $[0, T]$ into sufficiently many subintervals. Let us show (49).

$$\widetilde{\mathscr{I}}^\varepsilon_f(t) - \int_0^t f(\widehat{W}_r, r)\mathrm{d}W_r = (\mathcal{R}^\varepsilon(F^\varepsilon\Xi) - \mathcal{R}(F\Xi))(\mathbf{1}_{[0,t]})$$

$$= t\Big(\widehat{\Pi}^\varepsilon_0\Xi F^\varepsilon(0) - \Pi_0\Xi F(0))\Big)(t^{-1}\mathbf{1}_{[0,t]})$$

$$+ t\Big(\mathcal{R}^\varepsilon\Xi F^\varepsilon - \widehat{\Pi}^\varepsilon_0\Xi F^\varepsilon(0) - (\mathcal{R}\Xi F - \Pi_0\Xi F(0))\Big)(t^{-1}\mathbf{1}_{[0,t)}) .$$

We then obtain that the error is of order $\varepsilon^{\delta\kappa}$, $\delta \in (0, 1)$, using Theorem 3.14, Lemma 3.21 and (41) for the first term, and also Theorem 3.16 for the second term. Letting $\kappa \uparrow H$ and $M \uparrow \infty$, our total rate can be chosen arbitrary close to $H$.

To obtain the second estimate, we can bound $\widetilde{\mathscr{I}}^\varepsilon_f(t) - \widetilde{\mathscr{I}}^\varepsilon_{f,M}(t)$ with the first inequality in Theorem 3.16. □

## 3.3 | Nonconstant versus constant renormalization

If $\delta^\varepsilon$ comes from a mollifier (cf. Example 3.6), then the renormalization $\mathscr{C}^\varepsilon = \mathscr{K}^\varepsilon(\cdot, \cdot)$ that was applied in Theorem 3.14 and thus in Definition 3.24 is a constant, which is the familiar concept one encounters in the study of singular SPDE (Chandra & Hairer, 2016; Hairer, 2013, 2014). If $\delta^\varepsilon$ comes from wavelets such as the Haar basis, $\mathscr{K}^\varepsilon(\cdot, \cdot)$ is usually not constant but a periodic function with period $\varepsilon$. Thus, we see that our analysis gives rise to a "nonconstant renormalization." It is natural to ask if one can do with constant renormalization after all. Assume that $\mathscr{C}^\varepsilon$ is periodic with mean

$$C_\varepsilon = \frac{1}{\varepsilon}\int_0^\varepsilon \mathscr{C}^\varepsilon(t)dt.$$

From Lemma 3.13, it follows that $\mathscr{C}^\varepsilon$ (and its mean) are bounded by $\varepsilon^{H-1/2}$, uniformly in $t$. Putting all this together, it easily follows that $|\langle \mathscr{C}^\varepsilon - C_\varepsilon, \varphi\rangle| \lesssim \varepsilon^{\alpha+H-1/2}$, uniformly over all $\varphi$ bounded in $C^\alpha$, with convergence to zero when $\alpha > 1/2 - H$. As a consequence, taking $\varphi(t) = f(\widehat{W}^\varepsilon)$, for smooth $f$, we can clearly apply this estimate with any $\alpha < H$. Hence, by equating the constraints on $\alpha$, we arrive at $H > 1/4$. The practical consequence regarding part (i) of Theorem 3.25 then is that we can indeed replace nonconstant renormalization by constant renormalization, however at the prize of restricting to $H > 1/4$ and with an according loss on the convergence rate. Interestingly, our numerical simulation suggests that no loss occurs and constant renormalization works for any $H > 0$. While we have refrained from investigating this (technical) point further,[5] we can understand the mechanism at work by looking at the following toy example: Consider the Itô-integral $\int_0^1 W_t^H dW_t$ where $W^H$ is an fBm, but now with Hurst parameter $H > 1/2$, built, say, as Volterra process over $W$. Using Young integration theory, one can give a pathwise argument that shows that Riemann–Stieltjes approximations converge a.s. (with vanishing rate as $H \to 1/2$). However, we know from stochastic theory (Itô integration) that this convergence holds in $L^2$ (and then in probability) for any $H > 0$. We would thus expect that constant renormalization is still valid when $H \in (0, 1/4]$, but now the difference only vanishes in mean-square sense. This conjecture was checked numerically in Section 6.

## 3.4 | The case of the Haar basis

The following special case of the above approximations to $\int_0^t f(\widehat{W}_r, r)\mathrm{d}W_r$ is of particular interest for our purposes. We next collect some more concrete formulas that arise in this case.

Let $\varepsilon = 2^{-N}$, $\phi := \mathbf{1}_{[0,1)}$, and $\phi_{l,N} = 2^{N/2}\phi(2^N \cdot -l)$, $l \in \mathbb{Z}$. The corresponding approximation $\delta^\varepsilon$ to the Dirac delta is then

$$\delta^\varepsilon(x, y) = \sum_{l\in\mathbb{Z}} \phi_{l,N}(x)\phi_{l,N}(y) = 2^N \mathbf{1}_{[\lfloor x2^N \rfloor 2^{-N}, (\lfloor x2^N \rfloor+1)2^{-N})}(y), \quad x, y \in \mathbb{R}.$$

The mollified Volterra kernel (40) then takes the form

$$\begin{aligned}
\mathcal{K}^\varepsilon(u, v) &= \int_0^\infty \int_0^\infty \delta^\varepsilon(v, x_1)\delta^\varepsilon(x_1, x_2)K(u - x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \sqrt{2H} \cdot 2^N \int_{[\lfloor v2^N \rfloor 2^{-N}, (\lfloor v2^N \rfloor+1)2^{-N}\wedge u)} |u - x|^{H-1/2} \mathbf{1}_{\lfloor v2^N \rfloor 2^{-N} \le u} \,\mathrm{d}x \\
&= \frac{\sqrt{2H}}{1/2 + H} 2^N \times \\
&\quad \times \left(|u - \lfloor v2^N \rfloor 2^{-N}|^{1/2+H} - |u - (\lfloor v2^N \rfloor + 1)2^{-N} \wedge u)|^{1/2+H}\right) \mathbf{1}_{\lfloor v2^N \rfloor 2^{-N} \le u} .
\end{aligned}$$

A special role is played by the *diagonal function* as a renormalization,

$$\mathcal{C}^\varepsilon(t) = \mathcal{K}^\varepsilon(t, t) = \frac{\sqrt{2H}\, 2^N}{1/2 + H} |t - \lfloor t2^N \rfloor 2^{-N}|^{1/2+H} . \tag{51}$$

We additionally have

$$\begin{aligned}
\widehat{W}_t^\varepsilon &= \int_0^t K(t - r)\, \mathrm{d}W_r^\varepsilon = \sum_{l=0}^\infty Z_l \int_0^t K(t - r)\phi_{k,N}(r)\mathrm{d}r \\
&= \sum_{l=0}^\infty 2^{-N/2} \mathcal{K}^\varepsilon(t, l2^{-N}) Z_l = \sum_{l=0}^{\lfloor t2^N \rfloor} 2^{-N/2} \mathcal{K}^\varepsilon(t, l2^{-N}) Z_l ,
\end{aligned}$$

where $Z_l = \langle \dot{W}, \phi_{l,N} \rangle$ are i.i.d. $N(0,1)$-distributed variables. As our approximation, we can finally take $\mathscr{I}_f^\varepsilon(t)$ from Definition 3.24 with partition $\{[t_l, t_{l+1})\} = \{[l2^{-N}, (l+1)2^{-N} \wedge t)\}$, which gives us

$$\widetilde{\mathscr{I}}_{f,M}^{\varepsilon}(t) = \sum_{l=0}^{\lceil t2^N \rceil - 1} \sum_{m=0}^{M} \frac{1}{m!} \partial_1^m f(\widehat{W}_{t_l}^\varepsilon, t_l) 2^{N/2} Z_l \int_{t_l}^{t_{l+1}} \left( \widehat{W}_r^\varepsilon - \widehat{W}_{t_l}^\varepsilon \right)^m \mathrm{d}r - \\ - \sum_{m=1}^{M} \frac{1}{(m-1)!} \partial_1^m f(\widehat{W}_{t_l}^\varepsilon, t_{l+1}) \int_{t_l}^{t_{l+1}} \mathscr{C}^\varepsilon(r) \left( \widehat{W}_r^\varepsilon - \widehat{W}_{t_l}^\varepsilon \right)^{m-1} \mathrm{d}r$$

and

$$\widetilde{\mathscr{I}}_f^{\varepsilon}(t) = \sum_{l=0}^{\lceil t2^N \rceil - 1} \int_{t_l}^{t_{l+1}} [2^{N/2} Z_l \cdot f(\widehat{W}_r^\varepsilon, r)\, \mathrm{d}r - \mathscr{C}^\varepsilon(r)\, \partial_1 f(\widehat{W}_r^\varepsilon, r)]\, \mathrm{d}r.$$

As explained at the end of the last section, $\mathscr{C}^\varepsilon(r)$ in these formulas could be replaced by its local mean, the constant

$$2^N \int_0^{2^{-N}} \mathscr{C}^\varepsilon(r)\, \mathrm{d}r = \frac{\sqrt{2H}}{(H+1/2)(H+3/2)} 2^{N(1/2-H)}.$$

# 4 | THE FULL ROUGH VOLATILITY REGULARITY STRUCTURE

## 4.1 | Basic setup

We want to add an independent Brownian motion, so that we take an additional symbol $\overline{\Xi}$. We again fix $M$ and define

$$\overline{S} := S \cup \{\overline{\Xi}, \overline{\Xi}\mathcal{I}(\Xi), \dots, \overline{\Xi}\mathcal{I}(\Xi)^M\}, \quad \overline{\mathcal{T}} := \bigoplus_{\tau \in \overline{S}} \mathbb{R}\tau. \tag{52}$$

We fix $|\overline{\Xi}| = -1/2 - \kappa$ and the homogeneity of the other symbols is defined multiplicatively as before. We also set $\widehat{W}_t = \int_0^t K(s,t) dW_s$ with $K(s,t) = \sqrt{2H}|t-s|^{H-1/2}\mathbf{1}_{t>s}$, where $W$ and $\overline{W}$ are independent Brownian motions.

We extend the canonical model $(\Pi, \Gamma)$ to this regularity structure by defining

$$\Pi_s \overline{\Xi}\mathcal{I}(\Xi)^m := \left\{ t \mapsto \frac{\mathrm{d}}{\mathrm{d}t} \left( \int_s^t \left( \widehat{W}(u) - \widehat{W}(s) \right)^m d\overline{W}(u) \right) \right\}$$

(the above integral being in Itô sense), and[6]

$$\Gamma_{ts}\left( \overline{\Xi}\mathcal{I}(\Xi)^m \right) = \overline{\Xi}\Gamma_{ts}(\mathcal{I}(\Xi)^m).$$

Arguments similar to the proof of Lemma 3.9 show that this indeed defines a model on $\overline{\mathcal{T}}$.

## 4.2 | Small noise model large deviation

Given $\delta > 0$, we consider the "small-noise" model $(\Pi^\delta, \Gamma^\delta)$ on $\widetilde{T}$ obtained by replacing $W, \overline{W}$ by $\delta W, \delta\overline{W}$, which simply means that

$$\Pi^\delta \mathbf{1} = 1,$$

$$\Pi^\delta \mathcal{I}(\Xi)^m = \delta^m \Pi \mathcal{I}(\Xi)^m,$$

$$\Pi^\delta \Xi \mathcal{I}(\Xi)^m = \delta^{m+1} \Pi \Xi \mathcal{I}(\Xi)^m,$$

$$\Pi^\delta \overline{\Xi} \mathcal{I}(\Xi)^m = \delta^{m+1} \Pi \overline{\Xi} \mathcal{I}(\Xi)^m,$$

and

$$\Gamma^\delta_{ts} \mathbf{1} = \mathbf{1}, \Gamma^\delta_{ts} \Xi = \Xi, \Gamma^\delta_{ts} \overline{\Xi} = \overline{\Xi},$$

$$\Gamma^\delta_{ts} \mathcal{I}(\Xi) = \mathcal{I}(\Xi) + \delta(\widehat{W}(t) - \widehat{W}(s))\mathbf{1}$$

$$\Gamma^\delta_{ts} \tau\tau' = \Gamma^\delta_{ts} \tau \cdot \Gamma^\delta_{ts} \tau', \quad \text{for } \tau, \tau' \in \overline{S}.$$

Finally, for $h = (h_1, h_2)$ in $\mathcal{H} := L^2([0,T])^2$, we consider the deterministic model $(\Pi^h, \Gamma^h)$ defined by

$$\Pi^h \mathbf{1} = 1,$$

$$\Pi^h_s \Xi = h_1, \quad \Pi^h_s \overline{\Xi} = h_2,$$

$$\Pi^h_s \mathcal{I}(\Xi)(t) = \int_0^{t \vee s} (K(u,t) - K(u,s)) h_1(u) du,$$

$$\Pi^h \tau\tau' = \Pi^h \tau \Pi^h \tau' \text{ for } \tau, \tau' \in \overline{S}$$

and

$$\Gamma^h_{ts} \mathbf{1} = \mathbf{1}, \Gamma^h_{ts} \Xi = \Xi, \Gamma^h_{ts} \overline{\Xi} = \overline{\Xi},$$

$$\Gamma^h_{ts} \mathcal{I}(\Xi) = \mathcal{I}(\Xi) + \left(\int_0^{t \vee s} (K(u,t) - K(u,s)) h_1(u) du\right)\mathbf{1}$$

$$\Gamma^h_{ts} \tau\tau' = \Gamma^h_{ts} \tau \cdot \Gamma^h_{ts} \tau', \quad \text{for } \tau, \tau' \in \overline{S}.$$

The following lemma and theorem are proved in Appendix B.

**Lemma 4.1.** *For each $h \in \mathcal{H}$, $\Pi^h$ does define a model. In addition, the map $h \in \mathcal{H} \mapsto \Pi^h$ is continuous.*

**Theorem 4.2.** *The models $\Pi^\delta$ satisfy a LDP in the space of models with rate $\delta^2$ and rate function given by*

$$J(\Pi) := \begin{cases} \frac{1}{2}\|h\|^2_{\mathcal{H}} & \textit{if } \Pi = \Pi^h \textit{ for some } h \in \mathcal{H}, \\ +\infty, & \textit{otherwise}. \end{cases}$$

As an immediate corollary, we have the following corollary.

**Corollary 4.3.** *For $\delta$ small, $\mathbb{P}(Y_1^\delta \approx y) \approx \exp[-I(y)/\delta^2]$, in the precise sense of a large deviation principle (LDP) for*

$$Y_1^\delta := \int_0^1 f(\delta^H \widehat{W}_s)\delta\Big(\rho dW_s + \overline{\rho} d\overline{W}_s\Big)$$

*with speed $\delta^2$, and rate function given by*

$$I(y) := \inf_{h_1 \in L^2([0,1])} \left\{ \frac{1}{2}\|h_1\|_{L^2}^2 + \frac{\big(y - I_1(h_1)\big)^2}{2I_2(h_1)} \right\} \tag{53}$$

*where*

$$I_1(h_1) = \rho \int_0^1 f\left(\int_0^s K(u,s)h_1(u)du\right)h_1(s)ds, \quad I_2(h_1) = \int_0^1 f\left(\int_0^s K(u,s)h_1(u)du\right)^2 ds.$$

*Remark* 4.4. This improves a similar result obtained in Forde and Zhang (2017). In fact, we now cover functions $f$ of exponential form, as required in rough volatility modeling (Bayer et al., 2016, 2019; Gatheral et al., 2018).

*Proof.* Note that

$$Y_1^\delta = \left\langle \mathcal{R}^\delta F^\delta \cdot (\rho\Xi + \overline{\rho}\overline{\Xi}), 1_{[0,1]} \right\rangle,$$

where $F^\delta \equiv F^{\Pi^\delta}$ as defined in Lemma 3.21. By the contraction principle and the continuity estimate from Theorem 3.16, it holds that $Y_1^\delta$ satisfies an LDP, with rate function given by

$$I(y) = \inf\{\frac{1}{2}\Big(\|h_1\|_{L^2}^2 + \|h_2\|_{L^2}^2\Big), \quad y = \left\langle \mathcal{R}^h F^h \cdot (\rho\Xi + \overline{\rho}\overline{\Xi}), 1_{[0,1]} \right\rangle\},$$

where we used $F^h \equiv F^{\Pi^h}$. It then suffices to note that

$$\left\langle \mathcal{R}^h\Big(F^h \cdot (\rho\Xi + \overline{\rho}\overline{\Xi})\Big), 1_{[0,1]} \right\rangle = \int_0^1 f\left(\int_0^s K(u,s)h_1(u)du\right)\big(\rho h_1(s)ds + \overline{\rho}h_2(s)ds\big)$$

and optimizing over $h_2$ for fixed $h_1$ we obtain (53). □

We note that due to Brownian, respectively, fractional Brownian scaling, small-noise large deviations translate immediately to short-time large deviations, compare Forde and Zhang (2017).

Although the rate function here is not given in a very useful form, it is possible to expand it in small $y$ and so compute (explicitly in terms of the model parameters) higher order moderate deviations. In Bayer et al. (2019), this was related to implied volatility skew expansions.

## 5 | ROUGH VOLTERRA DYNAMICS FOR VOLATILITY

### 5.1 | Motivation from market micro-structure

Rosenbaum and coworkers, (El Euch et al., 2018; El Euch & Rosenbaum, 2019, 2018) show that stylized facts of modern market microstructure naturally give rise to fractional dynamics and leverage effects. Specifically, they construct a sequence of Hawkes processes, suitably rescaled in time and space, which converges in law to a rough volatility model of rough Heston form

$$dS_t/S_t = \sqrt{v_t}dB_t \equiv \sqrt{v}\Big(\rho dW_t + \overline{\rho}d\overline{W}_t\Big), \tag{54}$$

$$v_t = v_0 + \int_0^t \frac{a - bv_s}{(t-s)^{1/2-H}}ds + \int_0^t \frac{c\sqrt{v_s}}{(t-s)^{1/2-H}}dW_s.$$

(As earlier, $W, \overline{W}$ are independent Brownian motions.) Similar to the case of the classical Heston model, the square root provides both pain (with regard to any methods that rely on sufficient smooth coefficients) and comfort (an *affine structure*, here infinite-dimensional, which allows for closed-form computations of moment-generating functions). Arguably, there is no real financial reason for the square-root dynamics[7] and ongoing work attempts to modify the above square-root dynamics, such as to obtain (something close to) *log-normal volatility*. We note that log-normal volatility was a key feature of the rough volatility model discussed in Gatheral et al. (2018). This motivates the study of more general dynamic rough volatility models of the form

$$dS_t/S_t = f(Z_t)dB_t \equiv f(Z_t)\Big(\rho dW_t + \overline{\rho}d\overline{W}_t\Big)\,, \tag{55}$$

$$Z_t = z + \int_0^t K(s,t)v(Z_s)ds + \int_0^t K(s,t)u(Z_s)dW_s, \tag{56}$$

with sufficiently nice functions $f, u, v$. (While $f(x) = \sqrt{x}$ is a possible choice in what follows, we assume $u, v \in C^3$ for a local solution theory and then, in fact, impose $u, v \in C_b^3$ for global existence. One clearly expects nonexplosion under, for example, linear growth, but in order not to stray too far from our main line of investigation, we refrain from a discussion.) Remark that $f(z)$ plays the role of spot volatility. Further note that the choice $z = 0,\ v \equiv 0,\ u \equiv 1$ brings us back to the "simple" case with (rough stochastic) volatility $f(Z_t) = f(\widehat{W}_t)$ considered in earlier sections.

Equation (55) fits, with some good will, into the existing theory of stochastic Volterra equations with singular kernels (e.g., Pardoux & Protter, 1990 or Coutin & Decreusefond, 2001).[8]

## 5.2 | Regularity structure approach

We insist that (55) is not a classical Itô-SDE (solutions will not be semimartingales), nor a rough differential equation (in the sense of rough paths, driven by a Gaussian rough path as in P. K. Friz & Hairer, 2014, Chapter 10). If rough paths have established themselves as a powerful tool to analyze classical Itô-SDE, we here make the point that Hairer's theory is an equally powerful tool to analyze stochastic Volterra (respectively, mixed Itô-Volterra) equations in the singular regime of interest.

As preliminary step, we have to find the correct model space, spanned by symbols that arise by formal Picard iteration. To this end, rewrite (55) as an equation for modeled distributions,

$$\mathcal{Z} = \mathcal{I}(U(\mathcal{Z}) \cdot \Xi) + (\dots)\mathbf{1}, \tag{57}$$

from which one can guess (or rigorously derive along Hairer, 2014, Section 8.1) the need for the symbols

$$\mathbf{1}, \mathcal{I}(\Xi), \mathcal{I}(\Xi)^2, \mathcal{I}(\Xi\mathcal{I}(\Xi)), \dots$$

We have degrees $|\mathbf{1}| = 0$, $|\mathcal{I}(\Xi)| = H - \kappa$. For subsequent symbols, the degree is computed as

$$(1/2 + H) \times \{\text{number of } \mathcal{I}\} + (-1/2 - \kappa) \times \{\text{number of } \Xi\}.$$

For a modeled distribution, $\mathcal{Z}(t)$ takes values in the linear span of sufficiently many symbols, the (minimal) number of which is dictated by the Hurst parameter $H$. Loosely speaking, $\mathcal{Z} \in \mathcal{D}^\gamma$ indicates an expansions with $\gamma$-error estimate, in practice easy to see from the degree of the lowest degree symbols that do not figure in the expansion. For example, in case of a "level-2 expansion," we can expect

$$\mathcal{Z}(t) = (\dots)\mathbf{1} + (\dots)\mathcal{I}(\Xi) \in \mathcal{D}_0^{2(H-\kappa)}$$

as $|\mathcal{I}(\Xi)^2| = |\mathcal{I}(\Xi\mathcal{I}(\Xi))| = 2H - 2\kappa$. It follows from general theory (Hairer, 2014, Theorem 4.16) that if $\mathcal{Z} \in \mathcal{D}_0^\gamma$, then so is $U(\mathcal{Z})$, the composition with a smooth function, and by Hairer (2014, Theorem 4.7), the product with $\Xi \in \mathcal{D}^\infty_{-1/2-\kappa}$ is a modeled distribution in $\mathcal{D}^{\gamma-1/2-\kappa}$. For both reconstruction and convolution with singular kernels, one needs modeled distributions with positive degree $\gamma - 1/2 - \kappa > 0$. Given $H \in (0, 1/2]$, we can then determine which symbols (up to which degree) are required in the expansion. As earlier, fix an integer

$$M \geq \max\{m \in \mathbb{N} | m \cdot (H - \kappa) - 1/2 - \kappa \leq 0\}$$

(so that $(M + 1).(H - \kappa) - 1/2 - \kappa > 0$) and see that $\mathcal{Z} \in \mathcal{D}_0^{(M+1).(H-\kappa)}$ will do. When $H > 1/4$, and by choosing $\kappa > 0$ small enough, we see that $M = 1$ will do. That is, the symbols required to describe $\mathcal{Z}$ are $\{\mathbf{1}, \mathcal{I}(\Xi)\}$ and if one adds the symbols required to describe the right-hand side, one ends up with the level-2 model space spanned by

$$\{\Xi, \Xi\mathcal{I}(\Xi), \mathbf{1}, \mathcal{I}(\Xi)\},$$

which is exactly the model space for the "simple" rough pricing regularity structure, (28) in case $M = 1$. When $H \leq 1/4$, this precise correspondence is no longer true. To wit, in case $H \in (1/3, 1/4]$, taking $M = 2$ accordingly, solving (56) on the level of modeled distributions will require a ("level-3") model space given by

$$\langle \Xi, \Xi\mathcal{I}(\Xi), \Xi\mathcal{I}(\Xi)^2, \Xi\mathcal{I}(\Xi\mathcal{I}(\Xi)), \mathbf{1}, \mathcal{I}(\Xi), \mathcal{I}(\Xi)^2, \mathcal{I}(\Xi\mathcal{I}(\Xi))\rangle,$$

which is strictly larger than the corresponding level-3 simple model space given in (28). In general, one needs to consider an extended model space $\widehat{T} = \langle \widehat{S} \rangle$, so as to have, for any $m \geq 0$,

$$\tau_1, \dots, \tau_m \in \hat{S} \Rightarrow \Xi\mathcal{I}(\tau_1) \dots \mathcal{I}(\tau_m), \mathcal{I}(\tau_1) \dots \mathcal{I}(\tau_m) \in \hat{S},$$

(with the understanding that only finitely many such symbols are needed, depending on $H$ as explained above). As a result, symbols such as

$$\Xi\mathcal{I}(\Xi(\mathcal{I}(\Xi))^m), \ m \geq 0, \quad \mathcal{I}(\Xi(\mathcal{I}(\Xi(\mathcal{I}(\Xi))^m)^{m'}), m, m' \geq 0, \quad \dots$$

will appear. At this stage, a tree notation (omnipresent in the works of Hairer) would come in handy and we refer to Bruned, Chevyrev, Friz, amd Preiß (2019) (and the references therein) for a recent attempt to reconcile the tree formalism of branched rough path (Gubinelli, 2010; Hairer & Kelly, 2015) and the most recent algebraic formalism of regularity structures. (In a nutshell, the simple case (28) corresponds to trees where one node has $m$ branches; in the present nonsimple case, symbols branching can happen everywhere.) Carrying out the following construction in the general case of fixed $H > 0$

is certainly possible;[9] however, the algebraic complexity is essentially the one from branched rough paths, and hence, the general case requires a Hopf algebraic (Connes–Kreimer, Grossman–Larson, etc.) construction of the structure group (a.k.a. positive renormalization). Although this, and negative renormalization, is well understood (Bruned, Hairer, & Zambotti, 2019; Hairer, 2014, also Bruned et al., 2019, for a rough path perspective), all complete exposition would lead us to far astray from the main topic of this paper. Hence, for simplicity only, we shall restrict from here on to the level-2 case $H > 1/4$ (with $M = 1$ accordingly) but will mention general results whenever useful.

## 5.3 | Solving for rough volatility

We rewrite (56) as an equation for modeled distributions in $\mathcal{D}^\gamma$,

$$\mathcal{Z} = z\mathbf{1} + \mathcal{K}(U(\mathcal{Z}) \cdot \Xi + V(\mathcal{Z})). \tag{58}$$

(Here $U, V$ are the operators associated with composition with $u, v \in C^{M+2}$, respectively.) We also impose

$$\gamma \in (1/2 + \kappa, 1)$$

which is clearly necessary such as to have the product $U(\mathcal{Z}) \cdot \Xi$ in a modeled distribution space of positive parameter, so that reconstruction, convolution, and so on, make sense. Let $H > 1/4, M = 1$ and pick $\kappa \in (0, \frac{4H-1}{6})$ so that $(M + 1).(H - \kappa) - 1/2 - \kappa > 0$. As explained in the previous section, this exactly allows us to work in the familiar structure of Section 3.1. That is, with $M = 1$,

$$\mathcal{T} = \langle \Xi, \Xi\mathcal{I}(\Xi), \mathbf{1}, \mathcal{I}(\Xi) \rangle,$$

with index set and structure group as given in that section. This structure is equipped with the Itô-model and its (renormalization) approximations. Equation (58) critically involves the convolution operator $\mathcal{K}$ acting on $\mathcal{D}^\gamma$. The general construction (Hairer, 2014, Section 5) is among the most technical in Hairer's work, and, in fact, not directly applicable (our kernel $K$, although $\beta$-regularizing with $\beta = 1/2 + H$ fails the assumption 5.4 in Hairer, 2014), so we shall be rather explicit.

**Lemma 5.1.** *On the regularity structure* $(\mathcal{T}, A, G)$ *of Section 3.1 with* $M = 1$, *consider a model* $(\Pi, \Gamma)$ *that is* admissible *in the sense*

$$\Pi_t \mathcal{I}(\Xi) = (K * \Pi_t \Xi)(\cdot) - (K * \Pi_t \Xi)(t).$$

*Let* $\gamma > 0$, $F \in \mathcal{D}^\gamma$ *and set*[10]

$$\mathcal{K}F \, : \, s \in [0, T] \mapsto \mathcal{I}(F(s)) + (K * \mathcal{R}F)(s)\mathbf{1}.$$

*Then, (i)* $\mathcal{K}$ *maps* $\mathcal{D}^\gamma \to \mathcal{D}^{\min\{\gamma+\beta,1\}}$ *and (ii)* $\mathcal{R}(\mathcal{K}F) = K * \mathcal{R}F$, *that is, convolution commutes with reconstruction.*

*Remark* 5.2. Hairer (2014, Theorem 5.2) suggests the estimate $\mathcal{K}$ maps $\mathcal{D}^\gamma \to \mathcal{D}^{\gamma+\beta}$. The difference to our baby Schauder estimate stems from the fact, unlike assumption 5.3 in Hairer (2014, p. 64), we do not assume that our regularity structure contains the polynomial structure.

*Proof.* (Sketch) The special case $F \equiv \Xi \in \mathcal{D}^\infty$ was already treated in Lemma 3.19. We only show that, in the general case, $\mathcal{K}$ necessarily has the stated form but will not check the properties. It is enough to consider $F$ with values in $\langle \Xi, \Xi\mathcal{I}\Xi \rangle$ and make the ansatz

$$(\mathcal{K}F)(s) := \mathcal{I}F(s) + (\dots)\mathbf{1}\,.$$

Applying reconstruction, together with Hairer (2014, Proposition 3.28), we see that $\mathcal{R}(\mathcal{K}F) \equiv (\dots)$, which, in turn, must equal $K * \mathcal{R}F$, provided that we postulate validity of (ii). This is the given definition of $\mathcal{K}F$. □

We return to our goal of solving

$$\mathcal{Z} = z\mathbf{1} + \mathcal{K}(U(\mathcal{Z}) \cdot \Xi + V(\mathcal{Z})), \tag{59}$$

noting perhaps that $U(\mathcal{Z})$ makes sense for every function-like modeled distribution, say $F(t) = F_0(t)\mathbf{1} + \sum_{k=1}^{M} F_k(t)(\mathcal{I}\Xi)^k \in \mathcal{T}_+ := \langle \mathbf{1}, \mathcal{I}(\Xi), \dots, (\mathcal{I}\Xi)^M \rangle$, in which case

$$U(F)(t) = u(F_0(t))\mathbf{1} + u'(F_0(t)) \sum_{k=1}^{M} F_k(t)\mathcal{I}(\Xi)^k. \tag{60}$$

(Similar remarks apply to $V$, the composition operator associated with $v \in C^{M+2}$). Recall $M = 1$.

**Theorem 5.3.** *For any admissible model* $(\Pi, \Gamma)$ *and* $u,\ v \in C_b^{M+2}(\mathbb{R})$*, for any* $T > 0$*, Equation (59) has a unique solution in* $\mathcal{D}^\gamma(\mathcal{T}_+)$*, and the map* $(u, v, \Pi) \mapsto \mathcal{Z}$ *is locally Lipschitz in the sense that if* $\mathcal{Z}$ *and* $\widetilde{\mathcal{Z}}$ *are the solutions corresponding, respectively, to* $(u, v, \Pi)$ *and* $(\widetilde{u}, \widetilde{v}, \widetilde{\Pi})$*,*

$$|\!|\!|\mathcal{Z}; \widetilde{\mathcal{Z}}|\!|\!|_{\mathcal{D}_T^\gamma} \lesssim \|u - \widetilde{u}\|_{C_b^{M+2}} + \|v - \widetilde{v}\|_{C_b^{M+2}} + |\!|\!|(\Pi, \Gamma); (\widetilde{\Pi}, \widetilde{\Gamma})|\!|\!|_T\,,$$

*with the proportionality constant being bounded when the (respectively,* $C_b^{M+2}$ *and model) norms of the arguments stay bounded.*

*In addition, if* $(\Pi, \Gamma)$ *is the canonical Itô model (associated with Brownian, respectively, fBm,* $H > 1/4$*), then* $Z = \mathcal{R}\mathcal{Z}$ *solves (55) in the Itô sense.*

*Remark* 5.4. $Z = \mathcal{R}\mathcal{Z}$ is clearly the (unique) reconstruction of the (unique) solution to the abstract problem. We also checked that $Z$ is indeed a solution for the Itô-Volterra equation. However, if one desires to know that $Z$ is the unique strong solution to the stochastic Itô-Volterra equation, it is clear that one has to resort to uniqueness results of the stochastic theory, see, for example, Coutin and Decreusefond (2001).

*Proof.* The well-posedness and continuous dependence on the parameters essentially follows from results of Hairer (2014), and details are spelled out in Appendix C.

The fact that the reconstruction of the solution solves the Itô equation can be obtained by considering approximations, as is done in Hairer and Pardoux (2015, Theorem 6.2) or P. K. Friz and Hairer (2014, Chapter 5). □

Using the large deviation results obtained in the previous subsection, we can directly obtain an LDP for the log-price

$$X_t = \int_0^t f(Z_s)(\rho dW_s + \overline{\rho} d\overline{W}_s) - \frac{1}{2}\int_0^t f^2(Z_s)ds.$$

For square-integrable $h$, let $z^h$ be the unique solution to the integral equation

$$z^h(t) = z + \int_0^t K(s,t)u(z^h(s))h(s)ds.$$

**Corollary 5.5.** *Let $H \in (1/4, 1/2]$ and $f$ smooth (without boundedness assumption). Then $t^{H-\frac{1}{2}}X_t$ satisfies an LDP with speed $t^{2H}$ and rate function given by*

$$I(x) = \inf_{h\in L^2([0,1])} \left\{ \frac{1}{2}\|h\|_{L^2}^2 + \frac{\left(x - I_1^z(h)\right)^2}{2I_2^z(h)} \right\} \tag{61}$$

*where*

$$I_1^z(h) = \rho \int_0^1 f(z^h(s))h(s)ds, \quad I_2^z(h) = \int_0^1 f(z^h(s))^2 ds.$$

*Remark* 5.6. Concerning the case $H \le 1/4$, the following proof extends to any $H > 0$, *provided* that one builds the correct regularity structure as discussed at the end of Section 5.2. (In particular, the proof of Theorem 4.2 for obtaining Schilder-type large deviations for the appropriate Itô-model extends immediately.)

*Proof.* We ignore the second part $\int_0^t(\dots)ds$ in $X_t$, which is $O(t) = o(t^{\frac{1}{2}-H})$ as $f$ is bounded. Let $\hat{X}_t = \int_0^t f(Z_s)(\rho dW_s + \overline{\rho}d\overline{W}_s)$. By scaling, we see that $t^{H-\frac{1}{2}}\hat{X}_t$ is equal in law to $\hat{X}_1^\delta$, where $\delta = t^H$ and $X^\delta$, $Z^\delta$ are defined in the same way as $X$, $Z$ with $W, \overline{W}$ replaced by $\delta W, \delta\overline{W}$ and $v$ replaced by $v^\delta = \delta^{1+\frac{1}{2H}}h$. We then note that

$$X_1^\delta = \left\langle \mathcal{R}^\delta F(\mathcal{Z}^\delta)(\rho\Xi + \overline{\rho}\overline{\Xi}), 1_{[0,1]} \right\rangle =: \Psi(\Pi^\delta, v^\delta),$$

where $\Psi$ is locally Lipschitz by Theorem 5.3. We can then directly combine the fact that $\Pi^\delta$ satisfies an LDP (Theorem 4.2) with a contraction principle such as (Hairer and Weber, 2015, Lemma 3.3), to obtain that $X_1^\delta$ satisfies an LDP with rate function

$$I(x) = \inf\left\{ \frac{1}{2}(\|h\|_{L^2}^2 + \|\overline{h}\|_{L^2}^2, \quad x = \Psi(\Pi^{(h,\overline{h})}, 0) \right\}.$$

It then suffices to note that $z^h$ is exactly $\mathcal{R}\mathcal{Z}$ for $\mathcal{Z}$ the solution to (59) corresponding to a model $\Pi^{(h,\overline{h})}$ and with $h \equiv 0$, and to optimize separately over $\overline{h}$ as in the proof of Corollary 4.3. □

We also have an approximation result. (Assume $u, v$ to be smooth with three bounded derivatives.)

**Corollary 5.7.** *Let $H > 1/4$ (but see remark below). Then $Z = \lim Z^\varepsilon$, uniformly on compacts and in probability, where*

$$Z_t^\varepsilon = z + \int_0^t K(s,t)\big(u(Z_s^\varepsilon)dW_s^\varepsilon + (v(Z_s^\varepsilon) - \mathscr{C}^\varepsilon(s)uu'(Z_s^\varepsilon)ds\big). \tag{62}$$

*Remark* 5.8. Replacing the renormalization function $\mathscr{C}^\varepsilon$by its mean is possible, provided $H > 1/4$. However, unlike the discussion at the end of Section 3.2, this is no more a consequence of quantifying the distributional convergence. In the present context, this is achieved by checking directly model-convergence, which, fortunately, is not much harder. We leave details to the interested reader.

*Remark* 5.9. In contrast to the previous statement, the above result is more involved for $H \in (0, 1/4]$ and additional renormalization terms appear, the general description of which would benefit from pre-Lie products, as recently introduced Bruned et al. (2019).

*Proof.* Thanks to Theorem 3.14 and Theorem 5.3, it follows from continuity of reconstruction that

$$Z = \mathcal{R}\mathcal{Z} = \lim_{\varepsilon \to 0} \mathcal{R}^{\varepsilon} \mathcal{Z}^{\varepsilon},$$

so that the only thing to do is check that $Z^{\varepsilon}$ solves (62). Note that (59) implies that one has (omitting upper $\varepsilon$s at all normal and calligraphic $Z$ …)

$$\mathcal{Z}(t) = Z_t \mathbf{1} + u(Z_t)\mathcal{I}(\Xi),$$

and, with (60),

$$U(\mathcal{Z}(t))\Xi = u(Z_t)\Xi + u'u(Z_t)\mathcal{I}(\Xi)\Xi.$$

But then, because $\widehat{\Pi}^{\varepsilon}$ is a "smooth" model in the sense of remark 3.15 in Hairer (2014), one has

$$\begin{aligned}
\mathcal{R}^{\varepsilon}(U(\mathcal{Z}^{\varepsilon})\Xi)(t) &= \widehat{\Pi}^{\varepsilon}_t(U(\mathcal{Z}^{\varepsilon}(t))\Xi)(t) \\
&= u(Z^{\varepsilon}_t)(\widehat{\Pi}^{\varepsilon}_t \Xi)(t) + u' u(Z^{\varepsilon}_t)(\widehat{\Pi}^{\varepsilon}_t \Xi \mathcal{I}(\Xi))(t) \\
&= u(Z^{\varepsilon}_t)\dot{W}^{\varepsilon}(t) - u' u(Z^{\varepsilon}_t)\mathscr{K}^{\varepsilon}(t,t).
\end{aligned}$$

As convolution commutes with reconstruction, compare Lemma 5.1, it follows that $Z^{\varepsilon}$ is indeed a solution to (62). □

# 6 | NUMERICAL RESULTS

We will revisit the case of European option pricing under rough volatility. Building on the theoretical underpinnings of Section 3, we present a concise, but self-contained, description of the central algorithm of this paper—for simplicity restricted to the unit time interval—and complement the theoretical convergence rates obtained in previous sections with numerical counterparts. The code used to run the simulations has been made available on https://www.github.com/RoughStochVol.

## 6.1 | Implementation

Without loss of generality, set time to maturity $T = 1$. We are interested in pricing a European call option with spot $S_0$ and strike $K$ under rough volatility—the risk-free interest rate is assumed to satisfy $r = 0$. From Theorem 1.4, we have

$$C\big(S_0, K, 1\big) = \mathbb{E}\left[C_{BS}\left(S_0 \exp\left(\rho\mathscr{I} - \frac{\rho^2}{2}\mathscr{V}\right), K, \frac{\overline{\rho}^2}{2}\mathscr{V}\right)\right], \tag{63}$$

where the computational challenge obviously lies in the efficient simulation of

$$(\mathcal{I}, \mathcal{V}) = \left( \int_0^1 f(\widehat{W}_t, t)\mathrm{d}W_t, \int_0^1 f^2(\widehat{W}_t, t)\mathrm{d}t \right).$$

As explored in Subsection 3.4, we take a Wong–Zakai-style approach to simulating $\mathcal{I}$, that is, we approximate the white noise process $\dot{W}$ on the grid associated with the Haar basis as follows.

Let $\{Z_i\}_{i=1,\dots 2^N-1} \sim$ i.i.d. $\mathcal{N}(0,1)$ and choose a Haar grid level $N \in \mathbb{N}$ such that the step size of the grid satisfies $\varepsilon = 2^{-N}$. Then, for all $t \in [0,1]$, we set

$$\dot{W}^\varepsilon(t) := \sum_{i=0}^{2^N-1} Z_i e_i^\varepsilon(t), \quad \text{where} \quad e_i^\varepsilon(t) := 2^{N/2}\mathbf{1}_{[i2^{-N},(i+1)2^{-N})}(t), \tag{64}$$

which induces an approximation of the fBm

$$\widehat{W}^\varepsilon(t) = \sum_{i=0}^{2^N-1} Z_i \hat{e}_i^\varepsilon(t) \quad \text{with} \tag{65}$$

$$\hat{e}_i^\varepsilon(t) := \mathbf{1}_{t>i2^{-N}} \frac{\sqrt{2H}2^{N/2}}{H+1/2}\left(|t-i2^{-N}|^{H+1/2} - |t-\min((i+1)2^{-N},t)|^{H+1/2}\right). \tag{66}$$

As outlined above, the central issue is that the object $\int_0^1 f(\widehat{W}^\varepsilon(t),t)\dot{W}^\varepsilon(t)\mathrm{d}t$ does *not* converge in an appropriate sense to the object of interest $\mathcal{I}$ as $\varepsilon \to 0$. This is overcome by *renormalizing* the object, two possible approaches of which are explored in Subsection 3.4. For the remainder, we will consider the "simpler" renormalized object given by

$$\widetilde{\mathcal{I}}^\varepsilon = \int_0^1 f(\widehat{W}_t^\varepsilon, t)\dot{W}_t^\varepsilon \mathrm{d}t - \int_0^1 \mathscr{C}^\varepsilon(t)\partial_1 f(\widehat{W}_t^\varepsilon, t)\mathrm{d}t, \tag{67}$$

where the renormalization object $\mathscr{C}^\varepsilon(t)$ can be one of

$$\mathscr{C}^\varepsilon(t) = \begin{cases} \frac{2^N\sqrt{2H}}{H+1/2}|t - \lfloor t2^N \rfloor 2^{-N}|^{H+1/2}, \\ \frac{\sqrt{2H}}{(H+1/2)(H+3/2)}2^{N(1/2-H)}. \end{cases} \tag{68}$$

Inserting the nonconstant version of (68) into (67), we obtain

$$\widetilde{\mathcal{I}}^\varepsilon = \sum_{i=0}^{2^N-1} \int_{i2^{-N}}^{(i+1)2^{-N}} \left[ Z_i 2^{N/2} f(\widehat{W}^\varepsilon(t),t) - \frac{\sqrt{2H}2^N}{H+1/2}|t-i2^{-N}|^{H+1/2}\partial_1 f(\widehat{W}^\varepsilon(t),t)\right]\mathrm{d}t, \tag{69}$$

$$\mathscr{V}^\varepsilon := \sum_{i=0}^{2^N-1} \int_{i2^{-N}}^{(i+1)2^{-N}} f^2(\widehat{W}^\varepsilon(t),t)\mathrm{d}t. \tag{70}$$

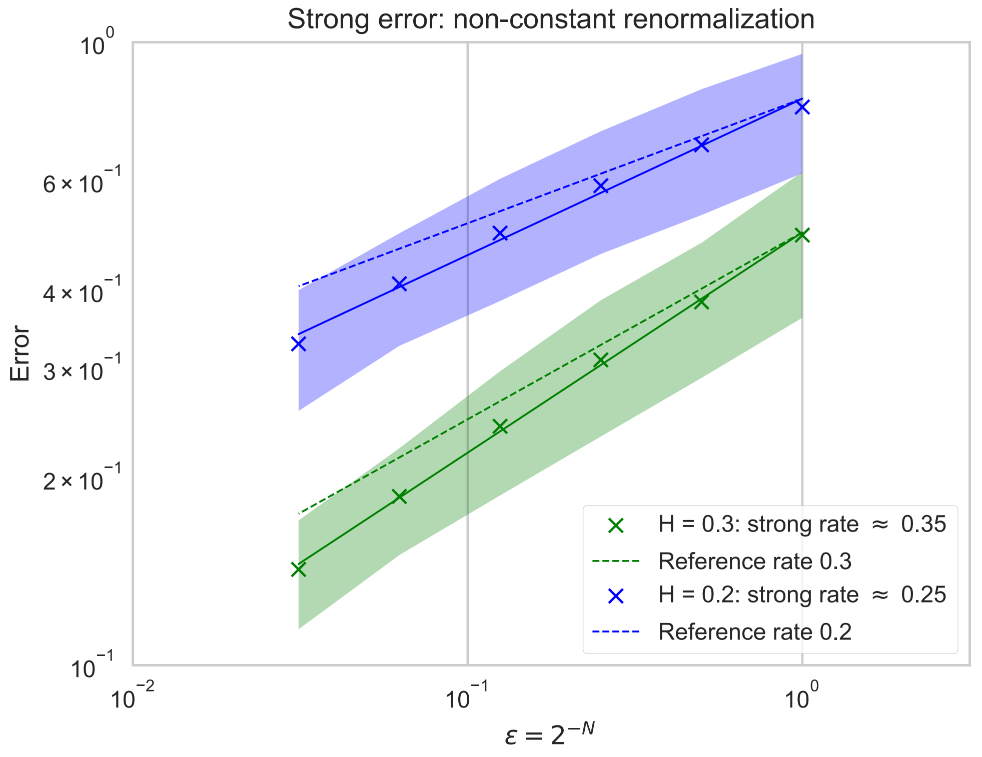


**FIGURE 1** Empirical strong errors in the sense of (71) on a log-log-scale under a *nonconstant* renormalization, obtained by $M = 10^5$ Monte Carlo samples with a trapezoidal rule delta of $\Delta = 2^{-17}$ and fineness of the reference Haar grid $\varepsilon' = 2^{-8}$ [Color figure can be viewed at wileyonlinelibrary.com]

*Note.* Solid lines visualize the empirical rates of convergence obtained by least squares regression, and dashed lines provide visual reference rates. Shaded color bands show interpolated 95% confidence regions based on normality of Monte Carlo estimator.

## 6.2 | Observed convergence rates

In this subsection, we will discuss strong convergence of the approximating object $\widetilde{\mathscr{I}}^\varepsilon$ to the actual object of interest $\mathscr{I}$ as well as weak convergence of the option price itself as the Haar grid interval size $\varepsilon \to 0$. Specifically, we will be looking at Monte Carlo estimates of our errors, that is, in order to approximate some quantity $\mathbb{E}[X]$ for some random variable $X$, we will instead be looking at $\frac{1}{M}\sum_{i=1}^M X_i$ where the $X_i$ are $M$ i.i.d. samples drawn from the same distribution as $X$. In other words, we need to generate $M$ realizations of the bivariate stochastic object $(\widetilde{\mathscr{I}}^\varepsilon, \mathscr{V}^\varepsilon)$, a task that can be vectorized as described below, thus avoiding expensive looping through realizations.

### 6.2.1 | Strong convergence

We verify Theorem 3.25(i) numerically, albeit in the $L^2(\Omega)$-sense and—for simplicity—with $f(x,t) := \exp(x)$, that is, with no explicit time dependence. That is, we are concerned with Monte Carlo approximations of

$$\left\| \widetilde{\mathscr{I}}^\varepsilon - \int_0^1 \exp(\widehat{W}_t)\mathrm{d}W_t \right\|_{L^2(\Omega)},$$

and we expect an error “almost” of order $H$.

**Algorithm 1:** Simulation of $M$ samples of $(\widetilde{\mathscr{I}}^{\varepsilon}, \mathscr{V}^{\varepsilon})$

**Parameters:** $M \in \mathbb{N}$: # Monte Carlo simulations
$N \in \mathbb{N}$: Haar grid 'level' such that $\varepsilon = 2^{-N}$
$d \in \mathbb{N}$: # discretization points of trapezoidal rule in each Haar subinterval

**Output:** $M$ samples of bivariate object $(\widetilde{\mathscr{I}}^{\varepsilon}, \mathscr{V}^{\varepsilon})$

initialize $\widetilde{\mathscr{I}}^{\varepsilon} = \mathscr{V}^{\varepsilon} = \mathbf{0} \in \mathbb{R}^M$;
simulate array $\mathbf{Z} \in \mathbb{R}^{M \times 2^N}$ of i.i.d. standard normals;
**for** each Haar subinterval $[i2^{-N}, (i+1)2^{-N})$ where $i \in \{0, \dots, 2^N - 1\}$ **do**
  choose discretization grid $\mathcal{D}^i$ with $d$ points on the Haar subinterval;
  evaluate functions $\widehat{e}_k^{\epsilon}, k = 0, \dots, i$, from (6.4) on $\mathcal{D}^i$ to obtain $\widehat{\mathbf{e}}^{\epsilon} \in \mathbb{R}^{(i+1) \times d}$;
  compute $\widehat{\mathbf{W}}^{\epsilon} = \mathbf{Z}^* \times \widehat{\mathbf{e}}^{\epsilon} \in \mathbb{R}^{M \times d}$ where $\mathbf{Z}^* \in \mathbb{R}^{M \times (i+1)}$ is the truncation of $\mathbf{Z}$ to its first $i+1$ columns such that $\widehat{\mathbf{W}}^{\epsilon}$ is an approximation of the fBm on $\mathcal{D}^i$;
  evaluate integrands from equations (6.7, 6.8) on $\mathcal{D}^i$ using $\widehat{\mathbf{W}}^{\epsilon}$ and the last column of $\mathbf{Z}^*$;
  approximate respective integrals on subinterval by trapezoidal rule ;
  add obtained estimates to running sums $\widetilde{\mathscr{I}}^{\varepsilon}$ and $\mathscr{V}^{\varepsilon}$;
**end**
**return** $\widetilde{\mathscr{I}}^{\varepsilon}, \mathscr{V}^{\varepsilon}$

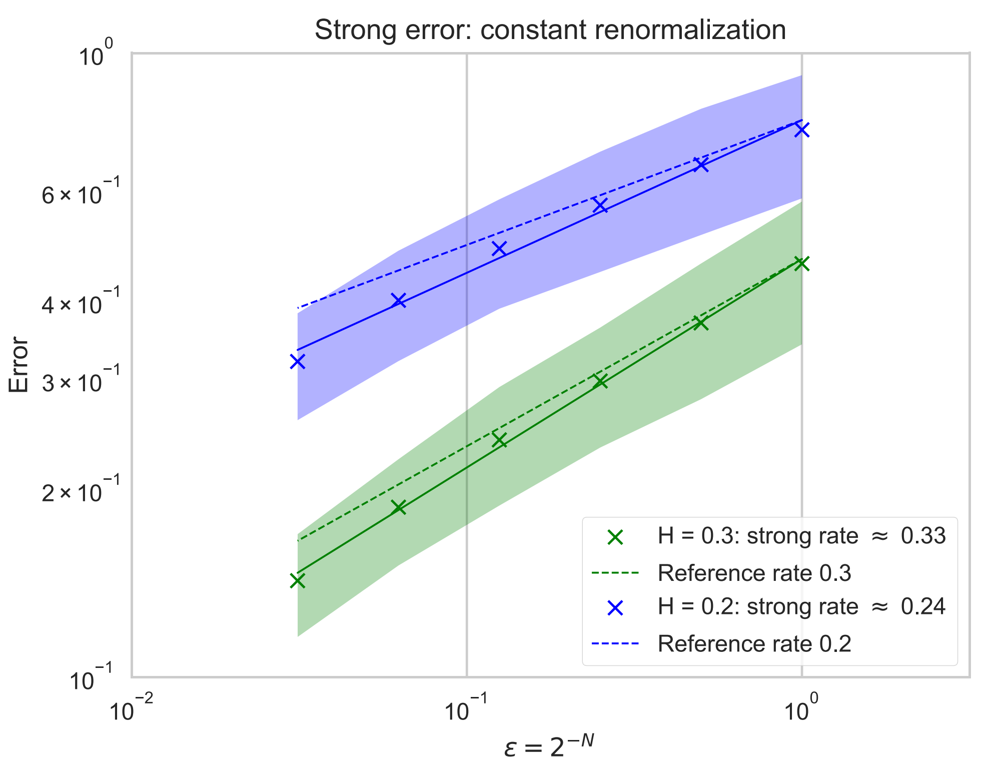


**FIGURE 2** Empirical strong errors in the sense of (71) on a log-log-scale under a *constant* renormalization, obtained by $M = 10^5$ Monte Carlo samples with a trapezoidal rule delta of $\Delta = 2^{-17}$ and fineness of the reference Haar grid $\varepsilon' = 2^{-8}$ [Color figure can be viewed at wileyonlinelibrary.com]

*Note.* Solid lines visualize the empirical rates of convergence obtained by least squares regression, and dashed lines provide visual reference rates. Shaded color bands show interpolated 95% confidence regions based on normality of Monte Carlo estimator.

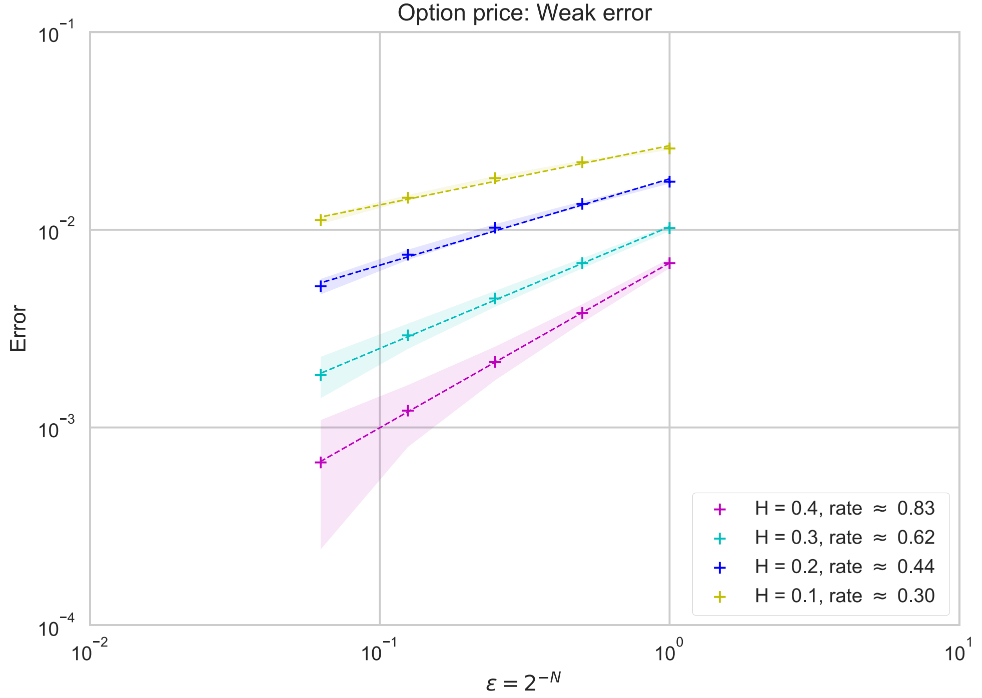


**FIGURE 3** Empirical weak errors in the sense of (74) on a log-log-scale as $\varepsilon \to \varepsilon' = 2^{-8}$, obtained with $M = 10^5$ MC samples with spot $S_0 = 1$, strike $K = 1$, correlation $\rho = -0.8$, spot vol $\sigma_0 = 0.2$, vvol $\eta = 2$, and trapezoidal rule delta $\Delta = 2^{-17}$ [Color figure can be viewed at wileyonlinelibrary.com]
*Note*. Dashed lines represent LS estimates for rate estimation, and shaded color bands show confidence regions based on normality of Monte Carlo estimator.

*Remark* 6.1. We choose $f(x,t) = \exp(x)$ because this closely resembles the *rough Bergomi* model (see Bayer et al., 2016, and below). Also, for the simplest nontrivial choice $f(x,t) = x$, the discretization error is overshadowed by the Monte Carlo error, even for very coarse grids.

As $(W, \widehat{W})$ is a two-dimensional Gaussian process with known covariance structure, it is possible to use the Cholesky algorithm (cf. Bayer et al., 2016, 2019) to simulate the joint paths on some grid, and then use standard Riemann sums to approximate the integral. The value obtained in this way could serve as a reference value for our scheme. However, for strong convergence, we need both objects to be based on the same stochastic sample. For this reason, we find it easier to construct a reference value by the wavelet-based scheme itself, that is, we simply pick some $\varepsilon' \ll \varepsilon$ and consider

$$\left\| \widetilde{\mathcal{F}}^{\varepsilon} - \widetilde{\mathcal{F}}^{\varepsilon'} \right\|_{L^2(\Omega)} \tag{71}$$

as $\varepsilon \to \varepsilon'$. As can be seen in Figures 1 and 2, both renormalization approaches stated in (68) are consistent with a theoretical strong rate of almost $H$ even for $H < 1/4$ (cf. discussion at the end of Section 3.2).

*Remark* 6.2 (Weak convergence). As the model does not have the Markov property, a proper weak convergence analysis proves to be subtle. Indeed, the rate of convergence of

$$\left|\mathbb{E}\left[\varphi\left(\widetilde{\mathscr{I}}^{\varepsilon}\right)\right]-\mathbb{E}\left[\varphi\left(\int_0^1 \exp(\widehat{W}_t)\mathrm{d}W_t\right)\right]\right|$$

as $\epsilon \to 0$ (for suitable test functions $\varphi$), remains an open problem. However, picking $\varphi(x) = x^2$, Itô's isometry yields

$$\mathbb{E}\left[\left(\int_0^1 \exp(\widehat{W}_t)\mathrm{d}W_t\right)^2\right] = \int_0^1 \mathbb{E}\left[\exp\left(2\widehat{W}_t\right)\right]\mathrm{d}t = \int_0^1 \exp\left(2t^{2H}\right)\mathrm{d}t, \tag{72}$$

which can be approximated numerically. So, we can consider

$$\left|\mathbb{E}\left[\left(\widetilde{\mathscr{I}}^{\varepsilon}\right)^2\right] - \int_0^1 \exp\left(2t^{2H}\right)\mathrm{d}t\right| \tag{73}$$

as $\varepsilon \to 0$. Our preliminary results indicate that for both renormalization approaches, the weak rate seems to be around the strong rate $H$.

### 6.2.2 | Option pricing

We pick a simplified version of the *rough Bergomi* model (Bayer et al., 2016), for which the instantaneous variance is given by

$$f^2(x) = \sigma_0^2 \exp(\eta x)$$

with $\sigma_0$ and $\eta$ denoting spot volatility and volatility of volatility, respectively. Let $C^{\varepsilon}$ denote the approximation of the call price (63) based on $(\widetilde{\mathscr{I}}^{\varepsilon}, \mathscr{V}^{\varepsilon})$, fix some $\varepsilon' \ll \varepsilon$, and consider

$$\left|C^{\varepsilon}\left(S_0, K, T=1\right) - C^{\varepsilon'}\left(S_0, K, T=1\right)\right| \tag{74}$$

as $\varepsilon \to \varepsilon'$. Empirical results displayed in Figure 3 indicate a weak rate of $2H$ across the full range of $0 < H < \frac{1}{2}$.

## ACKNOWLEDGMENTS

The authors acknowledge financial support from DFG research grants BA5484/1 (CB, BS) and FR2943/2 (PKF, BS), the ERC via Grant CoG-683164 (PKF), the ANR via Grant ANR-16-CE40-0020-01 (PG), and the DFG Research Training Group RTG 1845 (JM).

## ENDNOTES

[1] Other wavelet choices are possible. In particular, in the case of fractional noise, *Alpert-Rokhlin (AR) wavelets* have been suggested for improved numerical behavior; compare Grebenkov, Belyaev, and Jones (2016) where this is attributed to a series of works of A. Majda and coworkers. It would be desirable to explore this further in the context of rough volatility models.

[2] Hairer and Pardoux (2015) derive the KPZ result as special case of a Wong–Zakai result for Itô-SPDEs.

[3] In the section only, following P. K. Friz and Hairer (2014), symbols will be colored.

[4] $C^{|\Xi|}(\mathbb{R})$ denotes the space of distributions that are locally in the Besov space $\mathcal{B}^{|\Xi|}_{\infty,\infty}(\mathbb{R})$ (cf. Hairer, 2014, Remark 3.8).

[5] Some computations lead us to believe that this question can be settled with the aid of mixed $(1, \rho)$-variation of the covariance function of the Volterra process, compare P. K. Friz, Gess, Gulisashvili, and Riedel (2016), which we expect to hold uniformly over approximation.

[6] Upon setting $\Gamma_{ts}(\overline{\Xi}) = \overline{\Xi}$, the given relation is precisely implied by multiplicativity of $\Gamma$.

[7] This is also a frequent remark for the classical Heston model.

[8] We are not aware of any literature on mixed Itô-Volterra systems (although expect no difficulties). Here, of course, it suffices to first solves for $Z$ and then construct $S$ as stochastic exponential.

[9] We note that, as $H \downarrow 0$ the number of symbols tends to infinity. In comparison, among all recently studied singular SPDEs, only the sine-Gordon equation (Hairer & Shen, 2016) exhibits the similar feature of requiring arbitrarily many symbols.

[10] $\mathcal{I}$ is extended linearly to all of $\mathcal{T}$ by taking $\mathcal{I}\tau = 0$ for symbols $\tau \neq \Xi$.

## REFERENCES


Acworth, P. A., Broadie, M., & Glasserman, P. (1998). A comparison of some Monte Carlo and quasi Monte Carlo techniques for option pricing. In H. Niederreiter, P. Hellekalek, G. Larcher, & P. Zinterhof (Eds.), *Monte Carlo and quasi-Monte Carlo methods 1996* (pp. 1–18). New York, NY: Springer.

Alòs, E., León, J. A., & Vives, J. (2007). On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, *11*(4), 571–589.

Alòs, E., & Yang, Y. (2017). A fractional Heston model with $h > 1/2$. *Stochastics*, *89*(1), 384–399.

Avellaneda, M., Boyer-Olson, D., Busca, J., & Friz, P. (2003). Application of large deviation methods to the pricing of index options in finance. *Comptes Rendus Mathematique*, *336*(3), 263–266.

Bayer, C., Friz, P., & Gatheral, J. (2016). Pricing under rough volatility. *Quantitative Finance*, *16*(6), 887–904.

Bayer, C., Friz, P. K., Gulisashvili, A., Horvath, B., & Stemper, B. (2019). Short-time near-the-money skew in rough fractional volatility models. *Quantitative Finance*, *19*(5), 779–798.

Bayer, C., Friz, P. K., Riedel, S., & Schoenmakers, J. (2016). From rough path estimates to multilevel Monte Carlo. *SIAM Journal on Numerical Analysis*, *54*(3), 1449–1483.

Bayer, C., & Laurence, P. (2014). Asymptotics beats Monte Carlo: The case of correlated local vol baskets. *Communications on Pure and Applied Mathematics*, *67*(10), 1618–1657.

Bennedsen, M., Lunde, A., & Pakkanen, M. S. (2016). *Decoupling the short- and long-term behavior of stochastic volatility*. Preprint. arXiv:1610.00332.

Berestycki, H., Busca, J., & Florent, I. (2004). Computing the implied volatility in stochastic volatility models. *Communications on Pure and Applied Mathematics*, *57*(10), 1352–1373.

Bruned, Y., Chevyrev, I., Friz, P. K., & Preiß, R. (2019). A rough path perspective on renormalization. *Journal of Functional Analysis*, *277*(11), 108283. https://doi.org/10.1016/j.jfa.2019.108283

Bruned, Y., Hairer, M., & Zambotti, L. (2019). Algebraic renormalisation of regularity structures. *Inventiones mathematicae*, *215*(3), 1039–1156.

Chandra, A., & Hairer, M. (2016). *An analytic BPHZ theorem for regularity structures*. arXiv preprint arXiv:1612.08138.

Comte, F., Coutin, L., & Renault, É. (2012). Affine fractional stochastic volatility models. *Annals of Finance*, *8*(2–3), 337–378.

Comte, F., & Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, *8*(4), 291–323.

Coutin, L., & Decreusefond, L. (2001). *Stochastic Volterra equations with singular kernels* (pp. 39–50). Boston, MA: Birkhäuser.

Davis, M. H. A., & Mataix-Pastor, V. (2007). Negative Libor rates in the swap market model. *Finance and Stochastics*, *11*(2), 181–193.

Deuschel, J. D., Friz, P. K., Jacquier, A., & Violante, S. (2014a). Marginal density expansions for diffusions and stochastic volatility I: Theoretical foundations. *Communications on Pure and Applied Mathematics*, *67*(1), 40–82.

Deuschel, J. D., Friz, P. K., Jacquier, A., & Violante, S. (2014b). Marginal density expansions for diffusions and stochastic volatility II: Applications. *Communications on Pure and Applied Mathematics*, *67*(2), 321–350.

El Euch, O., Fukasawa, M., & Rosenbaum, M. (2018). The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, *22*(2), 241–280.

El Euch, O., & Rosenbaum, M. (2018). Perfect hedging in rough Heston models. *Annals of Applied Probability*, *28*(6), 3813–3856.

El Euch, O., & Rosenbaum, M. (2019). The characteristic function of rough Heston models. *Mathematical Finance*, *29*(1), 3–38.

Forde, M., & Zhang, H. (2017). Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, *8*(1), 114–145.

Friz, P., Gerhold, S., & Pinter, A. (2018). Option pricing in the moderate deviations regime. *Mathematical Finance*, *28*(3), 962–988.

Friz, P., & Riedel, S. (2011). Convergence rates for the full Brownian rough paths with applications to limit theorems for stochastic flows. *Bulletin des Sciences Mathématiques*, *135*(6), 613–628.

Friz, P. K., Gassiat, P., & Pigato, P. (2018). *Precise asymptotics: Robust stochastic volatility models*. arXiv preprint arXiv:1811.00267.

Friz, P. K., Gess, B., Gulisashvili, A., & Riedel, S. (2016). The Jain-Monrad criterion for rough paths and applications to random Fourier series and non-Markovian Hoermander theory. *Annals of Probability*, *44*(1), 684–738.

Friz, P. K., & Hairer, M. (2014). *A course on rough paths: With an introduction to regularity structures*. Cham: Springer International Publishing.

Fukasawa, M. (2011). Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance and Stochastics*, *15*(4), 635–654.

Fukasawa, M. (2017). Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, *17*(2), 189–198.

Gatheral, J., Jaisson, T., & Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, *18*(6), 933–949.

Gatheral, J., & Taleb, N. (2006). *The volatility surface: A practitioner's guide*. Hoboken, NJ: Wiley Finance.

Grebenkov, D. S., Belyaev, D., & Jones, P. W. (2016). A multiscale guide to Brownian motion. *Journal of Physics A: Mathematical and Theoretical*, *49*(4), 043001.

Gubinelli, M. (2010). Ramification of rough paths. *Journal of Differential Equations*, *248*(4), 693–721.

Gulisashvili, A. (2018). Large deviation principle for Volterra type fractional stochastic volatility models. *SIAM Journal on Financial Mathematics*, *9*(3), 1102–1136.

Hairer, M. (2013). Solving the KPZ equation. *Annals of Mathematics (2)*, *178*(2), 559–664.

Hairer, M. (2014). A theory of regularity structures. *Inventiones mathematicae*, *198*(2), 269–504.

Hairer, M. (2015). Introduction to regularity structures. *Brazilian Journal of Probability and Statistics*, *29*(2), 175–210.

Hairer, M., & Kelly, D. (2015). Geometric versus non-geometric rough paths. *Annales de l'Institut Henri Poincaré Probability and Statistics*, *51*(1), 207–251.

Hairer, M., & Pardoux, E. (2015). A Wong-Zakai theorem for stochastic PDEs. *Journal of the Mathematical Society of Japan*, *67*(4), 1551–1604.

Hairer, M., & Shen, H. (2016). The dynamical Sine-Gordon model. *Communications in Mathematical Physics*, *341*(3), 933–989.

Hairer, M., & Weber, H. (2015). Large deviations for white-noise driven, nonlinear stochastic PDEs in two and three dimensions. *Annales de la Faculté des Sciences de Toulouse. Mathématiques (6)*, *24*(1), 55–92.

Jacquier, A., Pakkanen, M. S., & Stone, H. (2018). Pathwise large deviations for the rough Bergomi model. *Journal of Applied Probability*, *55*(4), 1078–1092.

Janson, S. (1997). *Gaussian Hilbert spaces*. Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press.

Kusuoka, S. (2001). Approximation of expectation of diffusion processes and mathematical finance. In *Proceedings of Final Taniguchi Symposium, Nara 1998* (pp. 147–165). Advanced Studies in Pure Mathematics 31. Mathematical Society of Japan.

Lyons, T., & Victoir, N. (2004). Cubature on Wiener space. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *460*(2041), 169–198.

Meyer, Y., & Salinger, D. (1995). *Wavelets and operators*. Cambridge Studies in Advanced Mathematics, Vol. 1. Cambridge: Cambridge University Press.

Mijatović, A., & Tankov, P. (2016). A new look at short-term implied volatility in asset price models with jumps. *Mathematical Finance*, *26*(1), 149–183.

Neuenkirch, A., & Shalaiko, T. (2016). *The order barrier for strong approximation of rough volatility models*. arXiv preprint arXiv:1606.03854.

Ninomiya, S., & Victoir, N. (2008). Weak approximation of stochastic differential equations and application to derivative pricing. *Applied Mathematical Finance*, *15*(2), 107–121.

Nualart, D. (2006). Fractional Brownian motion: Stochastic calculus and applications. In *International Congress of Mathematicians*, Volume *3* (pp. 1541–1562). European Mathematical Society.

Nualart, D. (2013). *The Malliavin calculus and related topics*. New York, NY: Springer Science & Business Media.

Nualart, D., & Pardoux, É. (1988). Stochastic calculus with anticipating integrands. *Probability Theory and Related Fields*, *78*(4), 535–581.

Pardoux, E., & Protter, P. (1990). Stochastic Volterra equations with anticipating coefficients. *Annals of Probability*, *18*(4), 1635–1655.

Samko, S. G., Kilbas, A. A., & Marichev, O. I. (1993). *Fractional integrals and derivatives: Theory and applications*. Yverdon, Switzerland: Gordon and Breach Science Publishers. (translated from the 1987 Russian original, revised by the authors.)

**How to cite this article:** Bayer C, Friz PK, Gassiat P, Martin J, Stemper B. A regularity structure for rough volatility. *Mathematical Finance*. 2020;30:782–832. https://doi.org/10.1111/mafi.12233

## APPENDIX A: APPROXIMATION AND RENORMALIZATION (PROOFS)

**Lemma A.1.** *Given $a, b > 0$ and $\delta \in [0,1]$, we have for $x \notin [0,1)$*

$$|a^x - b^x| \le 2^{1-\delta}|x|^\delta (a^{x-\delta} \vee b^{x-\delta}) \cdot |a-b|^\delta,$$

*and for $x \in (0,1)$*

$$|a^x - b^x| \le 2^{1-\delta}|x|^\delta (a^{(x-1)\delta} b^{x(1-\delta)} \vee b^{(x-1)\delta} a^{x(1-\delta)}) \cdot |a-b|^\delta .$$

*Proof.* This follows from interpolation between $|a^x - b^x| \le |x| \sup_{z\in[a,b]} z^{x-1}|a-b| \le |x| a^{x-1} \vee b^{x-1}|a-b|$ and $|a^x - b^x| \le a^x + b^x \le 2a^x \vee b^x$. ☐

*Proof of Lemma* 3.8. Rewriting $\widehat{W}^\varepsilon(t) = \sqrt{2H}\int_0^\infty \mathrm{d}W(u) \int_0^\infty \mathrm{d}r\, \delta^\varepsilon(r,u)\, |t-r|^{H-1/2}\mathbf{1}_{r<t}$, we have

$$\begin{aligned}\mathbb{E}\left|\widehat{W}^\varepsilon(t) - \widehat{W}^\varepsilon(s)\right|^2 &= 2H\int_0^\infty \mathrm{d}u\left(\int_0^\infty \mathrm{d}r\delta^\varepsilon(r,u)(\mathbf{1}_{r<t}|t-r|^{H-1/2} - \mathbf{1}_{r<s}|s-r|^{H-1/2})\right)^2\\ &\lesssim \int_0^\infty \mathrm{d}u \int_0^\infty \mathrm{d}r|\delta^\varepsilon(r,u)|\big(\mathbf{1}_{r<t}|t-r|^{H-1/2} - \mathbf{1}_{r<s}|s-r|^{H-1/2}\big)^2\\ &\lesssim \int_0^{s\vee t} \mathrm{d}r\big(\mathbf{1}_{r<t}|t-r|^{H-1/2} - \mathbf{1}_{r<s}|s-r|^{H-1/2}\big)^2,\end{aligned}$$

where we used the Itô isometry in the first and Jensen's inequality in the second step. Assuming $s<t$, we can split the integral in domains $[0,s]$ and $[s,t]$ that yields the bound $|t-s|^{2H}\int_0^s |s-r|^{4H-1} + |t-s|^{2H} \lesssim |t-s|^{2H}$. Application of equivalence of moments for Gaussian random variables and Kolmogorov's criterion then shows the first inequality.

The second estimate follows by interpolation (and once more Kolmogorov), if we can prove that

$$\mathbb{E}|\widehat{W}^{\varepsilon}(t) - \widehat{W}(t)|^2 \lesssim \varepsilon^{2H-\kappa'}. \tag{A1}$$

We have, by Itô's isometry,

$$\mathbb{E}[\left|\widehat{W}^{\varepsilon}(t) - \widehat{W}(t)\right|^2] = 2H \int_0^{\infty} \mathrm{d}u \left( \int_0^{\infty} \mathrm{d}r\, \delta^{\varepsilon}(r,u)\, |t-r|^{H-1/2} \mathbf{1}_{r<t} - |t-u|^{H-1/2} \mathbf{1}_{u<t} \right)^2.$$

We can enlarge the inner integral such that $\int \delta^{\varepsilon}(r,u) dr = 1$ by neglecting an error term that can be estimated by $\int_{B(0,c\varepsilon)} \mathrm{d}u (\int_{B(0,c\varepsilon)} \mathrm{d}r\, \varepsilon^{-1} |t-r|^{H-1/2})^2 \lesssim \varepsilon^{2H}$. Application of Jensen's inequality then yields

$$\int_0^{\infty} \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}r\, |\delta^{\varepsilon}(r,u)| \left( |t-r|^{H-1/2} \mathbf{1}_{r<t} - |t-u|^{H-1/2} \mathbf{1}_{u<t} \right)^2.$$

The cases where either $r > u$ or $u > t$ yield an error term of order $\varepsilon^{2H}$. Hence, an application of Lemma A.1

$$\left| |t-r|^{H-1/2} - |t-u|^{H-1/2} \right| \lesssim (|t-r|^{-1/2+\kappa} + |t-u|^{-1/2+\kappa}) \cdot |u-r|^{H-\kappa}$$

finishes the proof of (A1). □

*Proof* (of (42)). We only consider the symbols $\Xi \mathcal{I}^m(\Xi)$, amd the symbols $\mathcal{I}(\Xi)^m$ can be handled with Lemma 3.8. In view of Lemmas 3.10 and 3.12, we have to control (for $m \geq 0$ in the first equation and $m > 0$ in the second equation)

$$\mathbb{E}\left| \int_0^{\infty} \mathrm{d}W^{\varepsilon}(t) \diamond \varphi_s^{\lambda}(t) (\widehat{W}_{st}^{\varepsilon})^m - \int_0^{\infty} \mathrm{d}W(t) \diamond \varphi_s^{\lambda}(t) (\widehat{W}_{st})^m \right|^2 \lesssim \varepsilon^{2\delta\kappa'} \lambda^{2mH-1-2\kappa'}, \tag{A2}$$

$$\mathbb{E}\left| \int_0^{\infty} \mathrm{d}t\, \varphi_s^{\lambda}(t) \Big( \mathscr{K}^{\varepsilon}(s,t) (\widehat{W}_{st}^{\varepsilon})^{m-1} - K(s-t) (\widehat{W}_{st})^{m-1} \Big) \right|^2 \lesssim \varepsilon^{2\delta\kappa'} \lambda^{2mH-1-2\kappa'}, \tag{A3}$$

where $\widehat{W}_{st}^{(\varepsilon)} = \widehat{W}^{(\varepsilon)}(t) - \widehat{W}^{(\varepsilon)}(s)$ and where $\delta \in (0,1)$, $\kappa' \in (0,H)$ is arbitrary. Equivalence of norms in the Wiener chaos and a version of Kolmogorov's criterion for models (Hairer, 2014, Proposition 3.32) then gives (42) (note that this gives a better homogeneity and then we actually need as we only subtract $2\kappa'$ and not $2m\kappa'$ in the exponent of $\lambda \in (0,1]$). We can rewrite the random variable in the expectation of (A2) using (Janson, 1997, Theorem 7.40) as

$$\int_0^{T+1} \mathrm{d}W(t) \diamond \int \mathrm{d}u\, \delta^{\varepsilon}(t,u) \Big( \mathbf{1}_{u\geq 0} \varphi_s^{\lambda}(u) (\widehat{W}_{su}^{\varepsilon})^m - \varphi_s^{\lambda}(t) (\widehat{W}_{st})^m \Big).$$

Using Janson (1997, Theorem 7.39) and Jensen's inequality, we can estimate the second moment of this Skorokhod integral by

$$\mathbb{E}|(A2)|^2 \lesssim \int_0^{T+1} \mathrm{d}t \int \mathrm{d}u\, |\delta^{\varepsilon}(t,u)|\, \mathbb{E}\Big( \mathbf{1}_{u\geq 0} \varphi_s^{\lambda}(u) (\widehat{W}_{su}^{\varepsilon})^m - \varphi_s^{\lambda}(t) (\widehat{W}_{st})^m \Big)^2.$$

In the regime $\lambda \leq \varepsilon$, every term in the parentheses can simply be bounded (using Lemma 3.8) by $\lambda^{2H-1} \lesssim \lambda^{2H-1-2\kappa'} \varepsilon^{\kappa'}$. If, on the other hand, $\varepsilon < \lambda$, we can split off a term of order

$\int_{B(0,c\varepsilon)} \mathrm{d}t \int_{B(0,c\varepsilon)} \frac{\mathrm{d}u}{\varepsilon} \lesssim \lambda^{2mH-1-2\kappa'}\varepsilon^{2\kappa'}$, drop the indicator $\mathbf{1}_{u\geq 0}$, and can bound on the support of $\delta^\varepsilon(t,u)$

$$|\varphi_s^\lambda(u)(\widehat{W}^\varepsilon_{su})^m - \varphi_s^\lambda(t)(\widehat{W}_{ts})^m| \leq |(\varphi_s^\lambda(u) - \varphi_s^\lambda(t)) \cdot |\widehat{W}^\varepsilon_{su}|^m + |\varphi_s^\lambda(t)| \cdot \left|(\widehat{W}^\varepsilon_{su})^m - (\widehat{W}_{st})^m\right|$$
$$\lesssim C_\varepsilon \mathbf{1}_{B(s,(1+2c)\lambda)}(t)\, \lambda^{-1-\kappa'}\varepsilon^{\kappa'}\lambda^{mH} + C_\varepsilon \mathbf{1}_{B(s,\lambda)}(t)\lambda^{-1}\lambda^{mH-\kappa'}\varepsilon^{\kappa'},$$

where $C_\varepsilon > 0$ denote random constants that are uniformly bounded in $L^p$ for $p \in [1,\infty)$. This shows (A2). To estimate (A3), we first note that due to $\mathbb{E}|(\widehat{W})^{m-1}_{st} - (\widehat{W}^\varepsilon_{st})^{m-1}|^2 \lesssim |t-s|^{2(m-1)H-2\kappa'}\varepsilon^{\delta 2\kappa'}$, we are only left with

$$\mathbb{E}\left|\int_0^\infty \mathrm{d}t\, \varphi_s^\lambda(t)(\mathscr{K}^\varepsilon(s,t) - K(s-t))(\widehat{W}^\varepsilon_{st})^{m-1}\right|^2 \lesssim \int_0^\infty \mathrm{d}t\, \varphi_s^\lambda(t)|\mathscr{K}^\varepsilon(s,t) - K(s-t)|^2\, |s-t|^{2(m-1)H},$$

which is straightforward to bound with Lemma 3.13 if $\lambda \leq \varepsilon$. For $\lambda < \varepsilon$ and $t > 2c\varepsilon$ with $c > 0$ as in Definition 3.5, the desired bound follows from Lemma A.2. The remaining case, however, contributes with

$$\int_{B(0,2c\varepsilon)} \mathrm{d}t\, \varphi_s^\lambda(t)|t-s|^{2(m-1)H}(\varepsilon^{2H-1} + |t-s|^{2H-1})$$
$$\lesssim \int_{B(s,\lambda^{-1}2c\varepsilon)} \mathrm{d}t\,(\lambda^{2(m-1)H}\varepsilon^{2H-1} + \lambda^{2mH-1}|t|^{2mH-1})$$
$$\lesssim \lambda^{2(m-1)H-1}\varepsilon^{2H} + \lambda^{2mH-1}(\lambda^{-1}\varepsilon)^{2mH} \leq \lambda^{2kH-\kappa'}\varepsilon^{\kappa'},$$

which completes the proof. □

**Lemma A.2.** *For $c$ as in Definition 3.5 and $t > 2c\varepsilon$ and $s \in \mathbb{R}$, we have for $\kappa' \in (0,H)$*

$$|K(s-t) - \mathscr{K}^\varepsilon(s,t)| \lesssim |s-t|^{H-1/2-\kappa'}\varepsilon^{\kappa'}.$$

*Proof.* If $2c\varepsilon \geq |s-t|/2$, the bound easily follows from Lemma 3.13. If $2c\varepsilon \geq |s-t|/2$, we can reshape

$$|K(s-t) - \mathscr{K}^\varepsilon(s,t)| = \left|\int_{-\infty}^\infty \mathrm{d}u\, \delta^{2,\varepsilon}(t,u)(\mathbf{1}_{t<s}|s-t|^{H-1/2} - \mathbf{1}_{s<u}|s-u|^{H-1/2})\right|,$$

where $\delta^{2,\varepsilon}(t,\cdot) := \int_{-\infty}^\infty \mathrm{d}x_1 \int_{-\infty}^\infty \mathrm{d}x_2 \delta^\varepsilon(t,x_1)\delta^\varepsilon(x_1,\cdot)$ satisfies the properties in Definition 3.5 with support in $B(t,2c\varepsilon)$. Note that for $2c\varepsilon \geq |s-t|/2$, either both indicator functions vanish or none so that we only have to consider $t < s$ where we obtain with Lemma A.1 up to a constant $\int_{-\infty}^\infty |\delta^{2,\varepsilon}(t,u)||t-s|^{H-1/2-\kappa'}\varepsilon^{\kappa'} \lesssim |t-s|^{H-1/2-\kappa'}\varepsilon^{\kappa'}$. □

*Proof of Lemma* 3.17. We restrict ourselves to the proof of (44), the other three inequalities follow by similar arguments. We fix a wavelet basis $\phi_y = \phi(\cdot - y)$, $y \in \mathbb{Z}$, $\psi_y^j = 2^{j/2}\psi(2^j(\cdot - y))$, $j \geq 0$, $y \in 2^{-j}\mathbb{Z}$ and use in the following the notation $\phi_y^j = 2^{j/2}\phi(2^j(\cdot - y))$, $j \geq 0$, $y \in 2^{-j}\mathbb{Z}$. Within this basis, we can express the $\mathcal{B}^\beta_{1,\infty}$ regularity of $\varphi$ by

$$\sum_{y\in\mathbb{Z}} |(\varphi,\phi_y)_{L^2}| + \sup_{j\geq 0} 2^{j\beta} \sum_{y\in 2^{-j}\mathbb{Z}} 2^{-dj/2}|(\varphi,\psi_y^j)_{L^2}| \lesssim \|\varphi\|_{\mathcal{B}^\beta_{1,\infty}}.$$

Without loss of generality, we can assume that $\lambda = 2^{-j_0}$ is dyadic, so that by scaling

$$\sum_{y\in 2^{-j_0}\mathbb{Z}} |(\varphi_s^\lambda, \phi_y^{j_0})_{L^2}| + \sup_{j\ge j_0} 2^{(j-j_0)\beta} \sum_{y\in 2^{-j}\mathbb{Z}} 2^{-(j-j_0)d/2} |(\varphi_s^\lambda, \psi_y^j)_{L^2}| \lesssim 2^{j_0 d/2} \|\varphi\|_{\mathcal{B}^\beta_{1,\infty}}. \tag{A4}$$

We can now rewrite

$$\begin{aligned}
&(\mathcal{R}F - \Pi_s F_s)(\varphi_s^\lambda) = \\
&\sum_{y\in 2^{-j_0}\mathbb{Z}} (\mathcal{R}F - \Pi_s F_s)(\phi_y^{j_0}) \cdot (\phi_y^{j_0}, \varphi_s^\lambda)_{L^2} + \sum_{j\ge j_0} \sum_{y\in 2^{-j}\mathbb{Z}} (\mathcal{R}F - \Pi_s F_s)(\psi_y^j) \cdot (\psi_y^j, \varphi_s^\lambda)_{L^2} \\
&= \sum_{y\in 2^{-j_0}\mathbb{Z}} (\mathcal{R}F - \Pi_y F_y)(\phi_y^{j_0})\,(\phi_y^{j_0}, \varphi_s^\lambda)_{L^2} + \sum_{y\in 2^{-j_0}\mathbb{Z}} \Pi_y(F_y - \Gamma_{ys} F_s)(\phi_y^{j_0})\,(\phi_y^{j_0}, \varphi_s^\lambda) \qquad \text{(A5)} \\
&+ \sum_{\substack{j\ge j_0,\\ y\in 2^{-j}\mathbb{Z}}} (\mathcal{R}F - \Pi_y F_y)(\psi_y^j)\,(\psi_y^j, \varphi_s^\lambda)_{L^2} + \sum_{\substack{j\ge j_0,\\ y\in 2^{-j}\mathbb{Z}}} \Pi_y(F_y - \Gamma_{ys} F_s)(\psi_y^j)\,(\psi_y^j, \varphi_s^\lambda)_{L^2}. \qquad \text{(A6)}
\end{aligned}$$

Only finitely many terms in (A5) contribute, which all can be bounded a constant times $2^{-j_0\gamma} = \lambda^\gamma$. Moreover,

$$\begin{aligned}
(A6) &\lesssim \sum_{j\ge j_0} 2^{-j\gamma} + \sum_{j\ge j_0} \sum_{A\ni\alpha<\gamma} 2^{-j\alpha} 2^{-(\gamma-\alpha)j_0} \sum_{y\in 2^{-j}\mathbb{Z}} 2^{jd/2} |(\varphi_s^\lambda, \psi_y^j)_{L^2}| \\
&\lesssim \sum_{j\ge j_0} 2^{-j\gamma} + 2^{-\gamma j_0} \sum_{A\ni\alpha<\gamma} \sum_{j\ge j_0} 2^{-(j-j_0)\alpha} 2^{-(j-j_0)\beta} \lesssim 2^{-j_0\gamma} = \lambda^\gamma,
\end{aligned}$$

where we used $\beta + \alpha > 0$, $\alpha \in A$ in the last line. □

*Proof of Lemma* 3.23. It easy to check, using Taylor's formula, that for scaled Haar wavelets $\varphi_s^\lambda$ and $\gamma \in (0, (M+1)H)$

$$\mathbb{E}\left[\left|\int \varphi_s^\lambda(t)\, f(\widehat{W}(t), t)\mathrm{d}W(t) - \Pi_s F\Xi(s)(\varphi_s^\lambda)\right|^2\right]^{1/2} \lesssim \lambda^{(\gamma-1/2-\kappa)}, \tag{A7}$$

uniformly for $s$ in compact sets. The same argument as in the proof of Lemma 3.17 then implies that (A7) actually holds for compactly supported smooth function $\varphi$ (or even compactly supported functions in $\mathcal{B}^\beta_{1,\infty}(\mathbb{R}^d)$). Proceeding now as in Hairer (2014), we choose test functions $\eta, \psi \in C_c^\infty$ with $\eta$ even and $\operatorname{supp}\eta \subseteq \mathrm{B}(0,1)$, $\int \eta(\mathrm{t})\,\mathrm{dt} = 1$. We then obtain for $\psi^\delta(s) = \langle \psi, \eta_s^\delta \rangle$

$$\begin{aligned}
&\mathbb{E}\left[|\mathcal{R}F\Xi(\psi^\delta) - \int \psi^\delta(t)\, f(\widehat{W}(t), t))\mathrm{d}W(t)|^2\right]^{1/2} \\
&= \mathbb{E}\left[\left|\int \mathrm{d}x\, \psi(x) \left(\mathcal{R}F\Xi(\eta_x^\delta) - \int \eta_x^\delta(t)\, f(\widehat{W}(t), t))\mathrm{d}W(t)\right)\right|^2\right]^{1/2} \\
&\lesssim \int \mathrm{d}x\, \psi^2(x)\, \delta^{\gamma-1/2-\kappa} \overset{\delta\to 0}{\to} 0,
\end{aligned}$$

where we included a term $\Pi_x \Xi F(x)$ in the second step. It remains to note that

$$\int \psi^\delta(t)\, f(\widehat{W}(t),t))\mathrm{d}W(t) \overset{\delta\to 0}{\to} \int \psi(t)\, f(\widehat{W}(t),t)\mathrm{d}W(t)$$

in $L^2(\mathbb{P})$ and further $\mathcal{R}F\Xi(\psi^\delta) \to \mathcal{R}F\Xi(\psi)$ a.s. and thus in $L^2(\mathbb{P})$. Putting everything together, we obtain

$$\mathbb{E}\left[|\mathcal{R}F\Xi(\psi) - \int \psi(t)\, f(\widehat{W}(t),t)\mathrm{d}W(t)|^2\right] = 0,$$

which implies the first statement. For the second identity, we proceed in the same way, but now use Lemma A.3. □

**Lemma A.3.** *For $F \in L^2(\mathbb{P} \times \mathrm{Leb})$, we have*

$$\mathbb{E}\left[\left|\int F(t)\mathrm{d}W^\varepsilon(t)\right|^2\right] \lesssim \int \mathbb{E}\left[|F(t)|^2\right]\mathrm{d}t.$$

*Proof.* As a consequence of Definition 3.5, we see that $\int |\delta^\varepsilon(x,y)\mathrm{d}x|$ is bounded uniformly in $\varepsilon$ and $y$. We can thus normalize $|\delta^\varepsilon(\cdot,r)|$ to a probability density and apply Itô's isometry and Jensen's inequality to obtain

$$\int F(t)\mathrm{d}W^\varepsilon(t) = \int_0^\infty \int_0^\infty \delta^\varepsilon(t,r)\, F(t)\mathrm{d}t\, \mathrm{d}W(r).$$

□

## APPENDIX B: LARGE DEVIATIONS PROOFS

*Proof of Lemma* 4.1. The fact that $\Pi^h$ satisfies the algebraic constraints is obvious, so we focus on the analytic ones. The Sobolev embedding $L^2 \subset C^{-1/2}$ yields that $\Pi\Xi$, $\Pi\overline{\Xi}$ satisfy the right bounds. Noting that (by, e.g., Samko, Kilbas, & Marichev, 1993, Section 3.1) $\|K * h\|_{C^H} \le C\|h\|_{C^{-1/2}}$ gives the bound for $\Pi\mathcal{I}(\Xi)^m$. Finally, we note that using Cauchy–Schwarz's inequality

$$\begin{aligned}
\left|\left\langle \Pi_t \Xi \mathcal{I}(\Xi)^m, \phi_x^\lambda \right\rangle\right| &= \left|\int h_1(s)\big(K * h_1(s) - K * h_1(t)\big)^m \phi_x^\lambda(s) ds\right| \\
&\le \left(\sup_{|s-t|\le\lambda} \left|K * h_1(s) - K * h_1(t)\right|\right)^m \|h_1\|_{L^2} \|\|\phi_x^\lambda\|_{L^2} \\
&\lesssim \lambda^{mH-1/2}.
\end{aligned}$$

The inequality for $\Pi\overline{\Xi}\mathcal{I}(\Xi)^m$ follows in the same way, and the bounds for $\Gamma$ also follow.

Continuity in $h$ is proved by similar arguments which we leave to the reader. □

*Proof of Theorem* 4.2. The theorem is a special case of results in Hairer and Weber (2015) for large deviations of Banach-valued Gaussian polynomials. Let us recall the setting.

Let $(B, \mathcal{H}, \mu)$ be an abstract Wiener space and let us call $\xi$ the associated $B$-valued Gaussian random variable, and $(e_i)$ an orthonormal basis of $\mathcal{H}$ with $e_i \in B^*$. For a multi-index $\alpha \in \mathbb{N}^{\mathbb{N}}$ with only finitely many nonzero entries, define $H_\alpha(\xi) = \Pi_{i\ge 0} H_{\alpha_i}(\langle \xi, e_i\rangle)$, where the $H_n$, $n \ge 0$ are the usual Hermite polynomials. For a given Banach space $E$, the homogeneous Wiener chaos $\mathcal{H}^{(k)}(E)$ is defined as the closure in $L^2(E, \mu)$ of the linear space generated by elements of the form

$$H_\alpha(\xi)y, \quad |\alpha| = k, \ y \in E.$$

Also, define the inhomogeneous Wiener chaos $\mathcal{H}^k(E) = \oplus_{i=0}^k \mathcal{H}^{(i)}(E)$. Finally, for $\Psi \in \mathcal{H}^{(k)}(E)$ and $h \in \mathcal{H}$, we define $\Psi^{hom}(h) = \int \Psi(\xi + h)\mu(d\xi)$, and for $\Psi = \sum_{i\le k} \Psi_i \in \mathcal{H}^k(E)$, we let $\Psi^{hom} = (\Psi_k)^{hom}$.

Now let $E = \oplus_{\tau\in\mathcal{W}} E_\tau$, where $\mathcal{W}$ is a finite set and each $E_\tau$ is a separable Banach space. Let $\Psi = \oplus_{\tau\in\mathcal{W}} \Psi_\tau$ be a random variable such that each $\Psi_\tau$ is in $\mathcal{H}^{K_\tau}(E_\tau)$. Letting $\Psi^\delta = \oplus_\tau \delta^{K_\tau}\Psi_\tau$, Theorem 3.5 in Hairer and Weber (2015) states that $\Psi^\delta$ satisfies an LDP with rate function given by

$$I(\Psi) = \inf\left\{1/2\|h\|_{\mathcal{H}}^2, \quad \Psi = \oplus_{\tau\in\mathcal{W}} \Psi_\tau^{hom}(h)\right\}.$$

In our case, we apply this result with $\mathcal{W} = \{\Xi\mathcal{I}(\Xi)^m, \overline{\Xi}\mathcal{I}(\Xi)^m, 0 \le m \le M\}$ and each $E_\tau$ is the closure of smooth functions $(t, s) \mapsto \Pi_t\tau(s)$ under the norms

$$\|\Pi\tau\| = \sup_{\lambda,t,\phi} \lambda^{-|\tau|}\left|\left\langle \Pi_t\tau, \phi_t^\lambda \right\rangle\right|.$$

To obtain Theorem 4.2, it suffices then to identify $(\Pi\tau)^{hom}(h)$, which is done in the following lemma. □

**Lemma B.1.** *For each $\tau \in \mathcal{W}$ and $h \in \mathcal{H}$, $(\Pi\tau)^{hom}(h) = \Pi^h\tau$.*

*Proof.* We prove it for $\tau = \Xi\mathcal{I}(\Xi)^m$, the other cases are similar. Note that $\Psi \mapsto \Psi^{hom}(h)$ is continuous from $\mathcal{H}^k$ to $\mathbb{R}$ for fixed $h$ (by an application of the Cameron–Martin formula), and so it is enough to prove that

$$\lim_{\varepsilon\to 0} \left(\widehat{\Pi}^\varepsilon \tau\right)^{hom}(h) = \Pi^h\tau, \tag{B1}$$

where $\widehat{\Pi}^\varepsilon$ corresponds to the (renormalized model) with piecewise linear approximation of $\xi$. For any test function $\varphi$, by definition, one has

$$\left\langle \Pi_t^\varepsilon\tau, \varphi \right\rangle = -\left\langle I^\varepsilon, \varphi' \right\rangle,$$

where

$$I^\varepsilon(s) = \int_t^s ((K * \xi^\varepsilon)(u) - (K * \xi^\varepsilon)(t))^m \xi^\varepsilon(u)du - C_\varepsilon R_m^\varepsilon,$$

where $R_m^\varepsilon$ is a renormalization term that is valued in the lower order chaos $\mathcal{H}^m$, so that, by definition, it does not play a role in the value of $(\Pi\tau)^{hom}$. Now note that if $\Phi$ is a Wiener polynomial whose leading order term is given by $\Pi_{i=1}^k \langle \xi, g_i\rangle$ (where the $g_i$ are in $\mathcal{H}$), then $\Phi^{hom}(h) = \Pi_{i=1}^k \langle h, g_i\rangle$. In our case, this means that

$$(I^\varepsilon)^{hom}(s) = \int_t^s \left((K * h_1^\varepsilon)(u) - (K * h_1^\varepsilon)(t)\right)^m h_1^\varepsilon(u)du,$$

where $h_1^\varepsilon = \rho^\varepsilon * h_1$. In other words, we have $(\widehat{\Pi}^\varepsilon\tau)^{hom} = \Pi_\tau^{h^\varepsilon}$, and by continuity of $h \mapsto \Pi^h$, we obtain (B1). □

## APPENDIX C: PROOFS OF SECTION 5

The proof of Theorem 5.3 follows from the estimates in the lemmas below, using the standard procedure of taking a time horizon $T$ small enough to obtain a contraction and then iterating the procedure. Note that due to the global boundedness of $u$ and $v$, the estimates are uniform in the starting point $z$, so that one obtains global existence (unlike the typical situation in SPDE where the theory only gives local in time existence).

By translating $u$ and $v$, we can assume w.l.o.g. that the initial condition is $z = 0$. Then, the solution will be an element of $\mathcal{D}^{\gamma}_{0,T}(\Gamma) := \{\ F \in \mathcal{D}^{\gamma}_{T}(\Gamma), \quad F(0) = 0.\}$.

**Lemma C.1.** *Let $F$ and $\widetilde{F}$ in $\mathcal{D}^{\gamma}_{0,T}(\mathcal{T})$ for the respective models $(\Pi, \Gamma)$ and $(\widetilde{\Pi}, \widetilde{\Gamma})$. For each $\gamma < 1$ and $T \in (0,1]$, one has*

$$|||\mathcal{K}F; \mathcal{K}\widetilde{F}|||_{\mathcal{D}^{\gamma}_{T}(\Gamma),\mathcal{D}^{\gamma}_{T}(\widetilde{\Gamma})} \lesssim T^{\eta} |||F; \widetilde{F}|||_{\mathcal{D}^{\gamma+|\Xi|}_{T}(\Gamma),\mathcal{D}^{\gamma+|\Xi|}_{T}(\widetilde{\Gamma})}$$

*for some $\eta > 0$. The constant of proportionality only depends on $\gamma$ and the norms of $(\Pi, \Gamma)$ and $(\widetilde{\Pi}, \widetilde{\Gamma})$.*

*Proof.* The choice $\gamma < 1$ avoids the appearance of polynomial terms, compare Hairer (2014, Section 5). Note that if $F$ belongs to $\mathcal{D}^{\gamma}_{0,T}$ so does $\mathcal{K}F$. As $K$ is a regularizing kernel of order $\beta := \frac{1}{2} + H$ in the sense of Hairer (2014), it follows along the lines of Hairer (2014, Section 5) that

$$|||\mathcal{K}F; \mathcal{K}\widetilde{F}|||_{\mathcal{D}^{\bar{\gamma}}_{T}(\Gamma),\mathcal{D}^{\bar{\gamma}}_{T}(\widetilde{\Gamma})} \lesssim |||F; \widetilde{F}|||_{\mathcal{D}^{\gamma+|\Xi|}_{T}(\Gamma),\mathcal{D}^{\gamma+|\Xi|}_{T}(\widetilde{\Gamma})},$$

where we pick $\bar{\gamma} \in (\gamma, 1)$ such that $\bar{\gamma} \le \gamma + |\Xi| + \beta = \gamma + H - \kappa$. On the other hand, it is clear from the definition of $|||\cdot;\cdot|||$ that, as $\mathcal{K}F$ and $\mathcal{K}\widetilde{F}$ vanish at $t = 0$, it holds that

$$|||\mathcal{K}F; \mathcal{K}\widetilde{F}|||_{\mathcal{D}^{\gamma}_{T}(\Gamma),\mathcal{D}^{\gamma}_{T}(\widetilde{\Gamma})} \lesssim T^{\eta} |||\mathcal{K}F; \mathcal{K}\widetilde{F}|||_{\mathcal{D}^{\bar{\gamma}}_{T}(\Gamma),\mathcal{D}^{\bar{\gamma}}_{T}(\widetilde{\Gamma})}$$

for $\eta = \bar{\gamma} - \gamma$. □

**Lemma C.2.** *Let $G$ (respectively, $\widetilde{G}$) be the composition operator corresponding to $g$ (respectively, $\widetilde{g}$) $\in C^{M+2}_{b}$. Then, one has*

$$|||G(F); \widetilde{G}(\widetilde{F})|||_{\mathcal{D}^{\gamma}_{T}(\Gamma),\mathcal{D}^{\gamma}_{T}(\widetilde{\Gamma})} \lesssim \|G - \widetilde{G}\|_{C^{M+2}} + |||F; \widetilde{F}|||_{\mathcal{D}^{\gamma}_{T}(\Gamma),\mathcal{D}^{\gamma}_{T}(\widetilde{\Gamma})}$$

*with a proportionality constant depending only on $\gamma$ and the norms of $(\Pi, \Gamma)$, $(\widetilde{\Pi}, \widetilde{\Gamma})$, $F$, $\widetilde{F}$, $g$, $\widetilde{g}$.*

*Proof.* This follows from the estimate in Hairer (2014, Theorem 4.16). The joint continuity is not stated there but is clear from the triangle inequality. □