On the roughness of paths and processes

by

Jose Luis Avilez

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Statistics

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In recent years, a significant amount of the stochastic volatility literature has focused on modelling the "roughness" or irregularity of the unobserved volatility time series and its effect on option pricing. In many models, roughness often takes the role of both the Hölder regularity of a trajectory as well as the covariance of the stochastic process. To extend the rough volatility literature, we contribute in two ways: (i) we extend the pathwise stochastic calculus to include integrators whose paths are rougher than the typical paths of Brownian motion, and (ii) we study two deep learning methods which allow us to determine the exact roughness of a given sample path.

For (i), we study the concept of p-variation of a continuous trajectory along a sequence of refining partitions, and the question of uniqueness when one changes the choice of refining partition sequence. We find a condition, termed p-roughness, which implies uniqueness of p-variation across a wide family of partition sequences. We then use this property to show that the rough pathwise Itô integral and the rough local time of a trajectory remain unchanged across this wide family of partition sequences.

For (ii), we study the roughness exponent of volatility under the risk-neutral and under the physical measure. Under the risk-neutral measure, we introduce a class of neural networks, called functional neural networks, which are able to learn interpolation schemes for the volatility surface. We use this method to show the recovery of stochastic volatility parameters on the rough Bergomi model. Furthermore, we consider the measurement of roughness under the physical measure for a class of models more general than fractional Brownian motion. We find that under this more general class, volatility roughness seems consistent with the roughness of volatility modelled as a fractional Brownian motion.

Acknowledgements

To my supervisor, Professor Alexander Schied, for his support and for persuading me to pursue my MMath degree at Waterloo. Alex introduced me to the exciting topics of rough volatility and model-free finance, which form the basis for this thesis. He was always happy to sit down and chart a path for this work and for my academic career. For that, I am incredibly grateful.

To my committee, Professor Yi Shen and Professor Adam Kolkiewicz, for offering their time to read this thesis and attend its defense.

To my undergraduate research advisors, Professor Nico Spronk and Professor Brian Forrest, who over the summer semester in 2019 planted the seeds of mathematical curiosity and mathematical maturity, which still bear fruit today.

To Mary Lou Dufton and Lisa Baxter, for ensuring everything in M3 runs smoothly; they have both made my life far easier.

To my friends: Ava, Chachi, Delaney, German, Gov, Mo, and Rea, for their everlasting loyalty, coolness, and for always indulging my taste for Cabernet Franc and soft cheeses. I promise to come back home frequently to visit all of you.

To my family, for their unconditional love and support.

And to you, reader, for joining me at the end of this year-long adventure.

Table of Contents

List of Figures

Chapter 1

Introduction

In recent years, there has been significant interest in the mathematical finance and probability theory literature to study paths whose regularity differs from the regularity of the paths of Brownian motion. This regularity property is usually referred to as "roughness". In finance, such paths appear when studying the covariance structure and Hölder regularity of latent volatility time series, as in [37]. One of the key difficulties that arises from rough models is the loss of the semimartingale property. As a result, developing an integration theory that can both measure and deal with roughness is necessary.

In this section, we provide extensive motivation for roughness, describe existing tools to tackle roughness, and motivate theoretical extensions which are explored in later chapters of this thesis.

1.1 Rough volatility

One of the game-changing discoveries in mathematical finance in the twentieth century was the Black-Scholes model, which provides a closed-form formula to price European options [11]. In the Black-Scholes universe, a stock trajectory is modelled as a geometric Brownian motion, a process with constant volatility; specifically, the model has the following semimartingale dynamics:

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

where μ is a constant drift, σ a constant volatility, and B_t a Brownian motion. However, any casual observer of stock markets would cast doubt to the assertion the volatility is

constant by looking at the price chart for any major index. A more sophisticated observer would even study the implied volatility surface for most assets and show a strict deviation from the constant volatility assumption. A solution to this model misspecification is to treat volatility as non-constant.

One of the most popular models that treats volatility as non-constant is Dupire's local volatility model; in it, volatility is treated as a deterministic function of the stock price and the time [24]. While such model is easy to fit and perfectly replicates a given implied volatility surface, it produces unrealistic stock price dynamics. Another popular approach is to treat volatility as a stochastic process, as is the case in the Heston volatility model. The Heston model treats volatility as an Itô diffusion [43], thus providing more reasonable dynamics than the Dupire's model, although the volatility surfaces it produces are not consistent with those observed in the market [35]. Other popular stochastic volatility models include the Hull-White model [47] and the SABR model [40]. Finding and describing reasonable stochastic volatility models remains an active and fruitful area of mathematical finance.

In all the models listed above—usually termed "conventional stochastic volatility models"—the sample paths of volatility have the same roughness as Brownian motion; namely, they are $\frac{1}{2} - \epsilon$ Hölder continuous for any $\epsilon > 0$. Gatheral, Jaisson, and Rosenbaum [37] propose that, in reality, the time series for spot volatility is much rougher than Brownian motion, often with Hölder regularity as low as 0.1. In particular, they propose that volatility is best described by as a model driven by fractional Brownian motion:

$$\sigma_t = \exp(X_t) \tag{1.1}$$

$$X_t = m + \nu \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^{(H)} \tag{1.2}$$

where ν, m, α are constants, and $W_t^{(H)}$ is a fractional Brownian motion with Hurst parameter $H \in (0, 1)$. A fractional Brownian motion is a centred continuous Gaussian process with stationary increments and covariance given by

$$\mathrm{Cov}(W_t^{(H)}, W_s^{(H)}) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right)$$

We note that fractional Brownian motion reduces to a standard Brownian motion when $H = \frac{1}{2}$; as such, fractional Brownian motion generalises standard Brownian motion. More interestingly, the sample paths of fractional Brownian motion are almost surely $H - \epsilon$ Hölder continuous for any $\epsilon > 0$. Moreover, fractional Brownian motion is a semi-martingale if and only if $H = \frac{1}{2}$. As a result, the model stipulated in Equation (1.1) is not a semi-martingale.

Such volatility models based on representations of fractional Brownian motion with $H < \frac{1}{2}$ are called rough volatility models.

The study of rough volatility models introduces several exciting challenges. The first challenge is concerned with non-semi-martingale integrators. For instance, the Bichteller-Dellacherie theorem [9, 22] states that a process is a valid Itô integrator if and only if it is a semi-martingale. Equation (1.2) already contains an inadmissible Itô integrator, which exemplifies the need for a robust rough integration theory. For completeness, we remark that, by appealing to the theory of Young integration [76], one may define an integral when the sum of the Hölder regularities of the integrand and integrator is greater than one; such is the case in Equation (1.2), so we are content with interpreting this as a pathwise Young integral. In Section 1.2, we introduce an elementary approach to study pathwise integration, and in Section 2 we discuss an extension of pathwise integration for rough trajectories.

A second interesting challenge is the determination of the parameter that governs the roughness of paths and processes: the Hurst parameter H. Computing H is especially important as it pertains to questions of calibration of stochastic volatility models. A fast and accurate calibration of these models is necessary to price derivatives in a manner consistent with the observed European option prices in the market. We provide a brief overview of methods used to compute H in Section 1.3 and introduce two methods to find H under the physical and risk-neutral measures in Section 3.

1.2 Pathwise stochastic calculus

In his seminal paper *Calcul d'Itô sans probabilités* [27], Föllmer proved the following path-by-path Itô formula:

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s \tag{1.3}$$

where $f \in C^2(\mathbb{R})$, and the trajectory $X \in C[0,T]$ admits an analytical property called pathwise quadratic variation along a sequence of partitions. To define pathwise quadratic variation, we begin with a sequence of refining partitions $\pi = (\pi^n)_{n=1}^{\infty}$ of $[0,T]$ whose mesh satisfies $|\pi^n| = \sup_{t \in \pi^n} |t' - t| \to 0$; here t' is the successor to t in the ordered set π^n. Then, we say that a continuous trajectory X admits quadratic variation along π provided that

$$\lim_{n \to \infty} \sum_{\substack{s \in \pi^n \\ s \leq t}} (X_{s'} - X_s)^2 \tag{1.4}$$

converges pointwise for each $t \in [0,T]$ to a continuous function. In case convergence holds, we denote the limit by $\langle X \rangle_t \in C[0,T]$.

With quadratic variation at hand, we may now give meaning to the right-hand side of Equation (1.3). It is readily apparent that if $\langle X \rangle_t$ exists, it must be non-decreasing and thus of bounded variation; this property makes it a valid Lebesgue-Stieltjes integrator, and we may interpret the second integral in 1.3 as the Lebesgue integral of $f''(X_t)$ against the positive Radon measure defined by $\mu([s,t]) = \langle X \rangle_t - \langle X \rangle_s$. The remaining term can be defined as a non-anticipating Riemann sum:

$$\int_0^t f'(X_s) dX_s = \lim_{n \to \infty} \sum_{\substack{s \in \pi^n \\ s \leq t}} f'(X_s)(X_{s'} - X_s) \tag{1.5}$$

Notice that this integral may fail to exist as a Lebesgue-Stieltjes integral, as X need not be of bounded variation. We remark that this integral depends on the choice of refining partition sequence π; this becomes clear in Example 2.2.

The key advantages of Föllmer's calculus are twofold: (i) it is strictly more elementary than the stochastic calculus for continuous semimartingales, and (ii) it admits a wider class of integrators than classical stochastic integration. To see point (i), we need only note that the missing ingredient to prove the correctness of Equation (1.3) is the statement of Taylor's formula applied to f, after which the proof strategy becomes apparent. The stochastic calculus for continuous semimartingales may span an entire graduate probability course which starts by constructing Brownian motion and concludes with a proof of the Bichteler-Dellacherie theorem. That entire journey requires sophisticated machinery far beyond the elementary calculus required for (1.3). To see point (ii), we note that the Bichteler-Dellacherie theorem [9, 22] implies that the admissible nontrivial integrators for Itô integration are local martingales, which have strictly increasing quadratic variation. This precludes fractional Brownian motion with Hurst parameter $H \neq \frac{1}{2}$ from being valid Itô integrators. The typical paths of "smooth" fractional Brownian motion (i.e. when $H > 0.5$) are, however, valid integrators for Föllmer's calculus, as they have vanishing quadratic variation along any refining partition whose mesh goes to zero.

Föllmer's approach is particularly attractive for robust finance. In mathematical finance, an integral of the form of (1.5) often denotes the value process of a portfolio. Usually, this requires treating X as an asset price, usually modelled by a semimartingale, which requires the introduction of a probability model $\mathbb{P}$. Statistically determining $\mathbb{P}$ is hard, so the modeller must often worry about model misspecification. The Föllmer integral dispels with this as the only required ingredient is a continuous trajectory which admits

continuous quadratic variation. In particular, the pathwise approach removes Knightian risk—the appearance of "unknown unknowns" that arise from misspecified models; see [29] for a detailed discussion.

As such, Föllmer's calculus has been used to recover the Black-Scholes formula and local volatility models [10], hedge path-dependent options [71, 19], recover Dupire's functional Itô's formula [17], develop a model-free CPPI/DPPI strategy [68], and study functionally-generated portfolios [49, 70]; this list is, of course, non-exhaustive.

1.2.1 Rougher trajectories

One thing Föllmer's calculus cannot do is integrate with respect to trajectories which are, in a sense, rougher than Brownian motion. A key class of examples is the usual paths of processes driven by fractional Brownian whose Hurst parameter satisfies $H < \frac{1}{2}$. The reason for this is that, for fractional Brownian motion, the following equality holds almost surely:

$$\left\langle B^{(H)} \right\rangle_t^{(p)} = \lim_{n\to\infty} \sum_{\substack{s\in\pi^n \\ s\leq t}} (B^{(H)}_{s'} - B^{(H)}_s)^p = \begin{cases} 0 & p > \frac{1}{H} \\ t\mathbb{E}\left(\left[B_1^{(H)}\right]^p\right) & p = \frac{1}{H} \\ \infty & p < \frac{1}{H} \end{cases}$$

Since the seminal paper *Volatility is rough* by Gatheral, Jaisson, and Rosenbaum [37], the field of stochastic volatility modelling has become interested in the mathematics of processes whose paths are more irregular than Brownian motion. The authors of this paper argue that physical paths of asset log-volatility processes share statistical properties with fractional Brownian motion with $H < \frac{1}{2}$, which gives rise to the idea that "volatility is rough".

To translate the idea of roughness to a path-by-path setting, we need to extend the definition of quadratic variation.

Definition 1.1. Let $X \in C[0,T]$ and let $\pi = (\pi^n)_{n=1}^\infty$ be a refining sequence of partitions for $[0,T]$. We say that X admits **p-variation along** π if the limit

$$\langle X \rangle_t^{(p)} = \lim_{n\to\infty} \sum_{\substack{s\in\pi^n \\ s\leq t}} (X_{s'} - X_s)^p$$

exists for each $t \in [0,T]$ and the limit $\langle X \rangle^{(p)}$ is continuous. In this case, we say $X \in V_p(\pi)$.

We remark that this is different from other uses of the term "p-variation", which sometimes refers to the quantity

$$\sup_{\pi} \sum_{\substack{s \in \pi^n \\ s \leq t}} (X_{s'} - X_s)^p$$

One way to stabilise Föllmer's integral when it encounters rough integrators is to introduce more regularity to the function f in (1.3) and exploit this extra regularity by extracting more information from its Taylor development. This idea is formally encoded in the theorem below.

Theorem 1.1. (Cont, Perkowski [18]) Let π be a refining sequence of partitions for $[0, T]$, p an even integer, $f \in C^p[0, T]$, and $X \in V_p(\pi)$. Then, we have the following Itô formula:

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s + \frac{1}{p!} \int_0^t f^{(p)}(X_s) d \langle X \rangle_s^{(p)} \tag{1.6}$$

where the first term on the right-hand is defined by the compensated Riemann sum:

$$\int_0^t f'(X_s) dX_s = \lim_{n \to \infty} \sum_{\substack{s \in \pi^n \\ s \leq t}} \sum_{k=1}^{p-1} \frac{f^{(k)}(X_s)}{k!} (X_{s'} - X_s)^k \tag{1.7}$$

The proof of this theorem is attained by writing $f(X_t) - f(X_0)$ as a telescoping sum over a partition $\pi = (\pi^n)_{n=1}^{\infty}$, applying Taylor's theorem to the increments of $f(X)$ along the partition, and sending $n \to \infty$; this directly mimics the proof of Equation (1.3).

When $p = 2$, Theorem 1.1 reduces to Equation (1.3). For $p > 2$, the left-Riemann sum of $f'(X)$ against the increments of X must now be compensated with the higher-order terms appearing in the Taylor development for f. This bears resemblance with the construction of the integral against a geometric rough path. In rough path theory, one enhances a Banach space-valued path X_t with a second order process $\mathbb{X}$, which satisfies the so-called Chen's relation. This second order process may then be used as a compensator in the definition of the rough path integral of a path Y_t against the rough path $(X_t, \mathbb{X}_{s,t})$, (see Chapter 4 of [31]). In particular, the integral in Equation (1.7) coincides with the rough path integral

$$\int_0^t f'(X_s) d(X, \mathbb{X}) \tag{1.8}$$

where

$$\mathbb{X}_{s,t} = \left(X_t - X_s, \frac{(X_t - X_s)^2}{2!}, \ldots, \frac{(X_t - X_s)^p}{p!} \right)$$

It is worth noting that for $p = 2$ it is precisely the existence of pathwise quadratic variation that negates the need for an integral compensator (see Section 5.3 of [31] for a discussion). Finding further connections between the rough Föllmer calculus and rough paths theory has intrinsic mathematical value. Furthermore, adapting the rough pathwise calculus for applications in finance would provide an additional tool to study the rough volatility phenomena introduced in Section 1.1. In Chapter 2 we study the question of uniqueness of pathwise p-variation, and suggest some financial interpretations for the rough pathwise integral.

1.3 Finding H

In *Volatility is rough* [37], the authors use a least squares method to elucidate the roughness of the paths of volatility. They assume they have discrete observations of the volatility process on a uniform time grid with mesh Δ—$\sigma_0, \sigma_\Delta, \sigma_{2\Delta}, \ldots, \sigma_{k\Delta}, \ldots$—where these observations occur on the time interval $[0, T]$. Setting $N = \lfloor T/\Delta \rfloor$ and $q \geq 0$, one may compute the empirical q-order log-moments of volatility:

$$m(q, \Delta) = \frac{1}{N} \sum_{k=1}^{N} |\log(\sigma_{k\Delta}) - \log(\sigma_{(k-1)\Delta})|^q$$

The above equation can be viewed as an estimator for

$$\mathbb{E}\left(|\log(\sigma_\Delta) - \log(\sigma_0)|^q\right)$$

One then assumes that the process σ_t is such that for some $s_q > 0$ and $b_q > 0$, we have

$$N^{qs_q} m(q, \Delta) \to b_q \tag{1.9}$$

as $\Delta \to 0$. This assumption is equivalent to saying that the paths of σ_t belong to the following subset of Besov space: $B^{s_q}_{q,\infty} \setminus B^{s_q+\epsilon}_{q,\infty}$, for $\epsilon > 0$. The parameter s_p is called the path smoothness. A theorem by Gladyshev states that the paths of fractional Brownian motion $B^{(H)}_t$ belong to $B^H_{p,\infty}$ almost surely [38]; that is, $s_p = H$ for all $p \geq 0$. This is often referred to as the monofractal scaling property of fractional Brownian motion.

Taking logarithms in Equation (1.9) and using $N = \lfloor T/\Delta \rfloor$, we find that

$$\log m(q, \Delta) \sim \log b_q + q s_q \log \Delta$$

Thus, regressing $\log m(q, \Delta)$ against $\log \Delta$ for different values of q, we determine the behaviour of s_q. Specifically, if $\Delta = (\Delta_k)_{k=1}^M$ is a family of meshes and if one assumes $s_p = H$, then the least-squares estimator is given by:

$$\hat{H} = \frac{1}{p} \frac{\sum_{k=1}^M \left(\log \Delta_k - \overline{\log \Delta}\right) \left(\log m(p, \Delta_k) - \overline{\log m(p, \Delta)}\right)}{\sum_{k=1}^M (\log \Delta_k - \overline{\log \Delta})^2} \tag{1.10}$$

Using this estimator, Gatheral et al. [37] find that $s_q = H \approx 0.1$ for a wide variety of asset classes.

Several other methods have been proposed to study the Hurst exponent, including quasi-likelihood [32], Malliavin calculus [3], absolute moments [33], and machine learning [45, 72, 66] methods. Some of these methods estimate H from the physical trajectories of asset prices, whilst others do so by using option price data.

In Chapter 3, we extend two deep learning methods to estimate the roughness of stochastic processes, one under the physical measure and one under the risk-neutral measure.

Chapter 2

Uniqueness of pathwise p-variation

The objective of this chapter is to obtain an extension of a roughness property discovered by Cont and Das in [16]. In this paper, the authors prove a theorem of the following form: "if a continuous trajectory X admits quadratic variation $\langle X \rangle$ and satisfies a technical roughness property, then its quadratic variation along all sufficiently regular partitions π equals the unique limit $\langle X \rangle$". In their paper, Cont and Das name the technical property "quadratic roughness".

In this chapter, we revisit the concept of quadratic roughness, propose an extension from 2 to p-roughness, and provide an intrinsic definition of p-variation for functions which satisfy the p-roughness condition. Most of the extensions here arise from judiciously changing the exponents from 2 to p in the paper by Cont and Das. In particular, in this section we try to re-prove most of the theorems from the original quadratic roughness paper in the general p-roughness setting.

As we deal with quadratic variation along different partitions, we shall denote by $\langle X \rangle_{\pi}^{(p)}$ the p-variation of a continuous trajectory along π. Since this is a continuous function, we may evaluate it at any point in time $t \in [0, T]$, which we shall denote by $\langle X \rangle_{\pi}^{(p)}(t)$.

2.1 Balanced partitions

When defining the pathwise Itô integral, we examined the limit as the mesh—the length of largest subdivision in a partition—was sent to zero. It is natural to demand that the smallest subdivision is not too far away in length from the mesh.

Definition 2.1. (Balanced sequence of partitions) Let (π^n) be a sequence of partitions of $[0,T]$ with $\pi = \pi^n = (0 = t_0^n < t_1^n < \ldots < t_{N(\pi^n)}^n = T)$ and define

$$\underline{\pi^n} = \inf_{0 \le i \le N(\pi^n)-1} |t_{i+1}^n - t_i^n|$$

We say that $\pi = (\pi^n)_{n=1}^\infty$ is **balanced** if there exists a number $c > 0$ such that for all $n \ge 1$,

$$\frac{|\pi^n|}{\underline{\pi^n}} \le c$$

The set of all balanced partition sequences of $[0,T]$ is denoted by $\mathbb{B}[0,T]$.

While it is not required in the definition of balanced partition sequence, we can and will assume that all our partitions have vanishing mesh; i.e. $\lim_{n\to\infty} |\pi^n| = 0$.

Note that a fixed partition π^n is simply a finite subset of $[0,T]$; implicit in the above definition, we have that the cardinality of π^n is denoted by $N(\pi^n)$. The family of finite subsets of a set can be directed by inclusion. Thus, adding points to a partition is termed "refinement", while removing them (which can be thought of as sampling from a finite set) is called "coarsening". Often, we wish to control the rate at which we can sample from a partition sequence.

Suppose we have a parent partition sequence $\pi = (\pi^n)_{n=1}^\infty$ from which we want to derive another partition sequence $A = (A^n)_{n=1}^\infty$, such that $A^n \subseteq \pi^n$ for all n. To aid us in deriving this child sequence, we introduce a sampling function $p : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Here, $p(n,k)$ will indicate the k-th sampled point in the n-th partition, so that in particular $1 \le k \le N(A^n)$, where $N(A^n)$ indicates the number of points sampled in the child partition.

Definition 2.2. (Coarsening a balanced partition sequence) Let $\pi = (\pi^n)_{n=1}^\infty$ be a balanced partition sequence of $[0,T]$ with $|\pi^n| \to 0$. Fix $0 < \beta < 1$. A sequence of partitions $A = (A^n)_{n=1}^\infty$ is called a β-coarsening of π provided that A is a subpartition sequence of π, i.e. each A^n is a sub-sample of π^n, or in symbols:

$$A^n = (0 = t_{p(n,0)}^n < t_{p(n,1)}^n < \ldots < t_{p(n,N(A^n))}^n = T)$$

and, additionally, A is a balanced partition of $[0,T]$, and $|A^n|$ and $|\pi^n|^\beta$ are asymptotically comparable:

$$\limsup_{n\to\infty} \frac{|A^n|}{|\pi^n|^\beta} < \infty \qquad \limsup_{n\to\infty} \frac{|\pi^n|^\beta}{|A^n|} < \infty$$

Note that if a continuous path has p-th variation along a refining sequence of partitions, it has the same p-th variation along any subsequence of partitions; this is simply a consequence of the elementary fact that if a sequence converges, then all its subsequences converge to the same value.

To prove the main results of this chapter, we will need to perform several estimates that are afforded to us via the regularity of the paths we study, as well as the asymptotic behaviour of the partitions we focus on. For a quick reference, we conclude this section by listing and proving some properties of balanced partitions, which will aid us in making the necessary estimates in the next section.

Proposition 2.1. Let $\pi = (\pi_n)_{n=1}^{\infty}$ be a balanced and refining sequence of partitions of $[0, T]$. Then,

1. $|\pi^n| \leq c\underline{\pi^n} \leq c\frac{T}{N(\pi^n)}$
2. $\liminf_{n\to\infty} N(\pi^n)\underline{\pi^n} > 0$ and $\limsup_{n\to\infty} N(\pi^n)|\pi^n| < \infty$
3. $\limsup_{n\to\infty} \frac{N(\pi^{n+1})}{N(\pi^n)} < \infty \iff \limsup_{n\to\infty} \frac{|\pi^n|}{|\pi^{n+1}|} < \infty \iff \limsup_{n\to\infty} \frac{\pi^n}{\underline{\pi^{n+1}}} < \infty$

Proof. 1. The first inequality is the definition of balanced partitions. For the second inequality, it suffices to observe that $\underline{\pi^n} N(\pi^n) \leq T$.

2. From the inequality we just proved, it follows that for all $n \geq 1$ we have

$$N(\pi^n)\underline{\pi^n} \leq T \leq N(\pi^n)|\pi^n|$$

Further using the balanced property:

$$\liminf_{n\to\infty} N(\pi^n)\underline{\pi^n} = \liminf_{n\to\infty} N(\pi^n)|\pi^n|\frac{\underline{\pi^n}}{|\pi^n|} \geq \frac{1}{c}\liminf_{n\to\infty} N(\pi^n)|\pi^n| \geq \frac{T}{c} > 0$$

Similarly,

$$\limsup_{n\to\infty} N(\pi^n)|\pi^n| = c\limsup_{n\to\infty} N(\pi^n)\underline{\pi^n} \leq cT < \infty$$

3. We can augment the inequality in (1) by further using the balanced property to obtain

$$kN(\pi^n)|\pi^n| \leq N(\pi^n)\underline{\pi^n} \leq T \leq N(\pi^n)|\pi^n| \leq k'N(\pi^n)\underline{\pi^n}$$

Where k, k' are suitable constants that come from the definition of balanced partitions. From these inequalities, the equivalence follows.

□

When comparing two different balanced partitions, we will often want to control their relative asymptotic behaviour. This is especially important when showing that, under certain conditions, a path's quadratic variation remains invariant under the choice of balanced partition sequences.

Lemma 2.1. Let $\pi = (\pi^n)_{n=1}^{\infty}$ and $\tau = (\tau^n)_{n=1}^{\infty}$ be balanced and refining partition sequences for $[0,T]$, such that $\limsup_{n\to\infty} \frac{|\pi^n|}{|\tau^n|} < 1$. Then, there exists a subsequence $(\pi^{k(n)})_{n=1}^{\infty}$ of π such that

$$\limsup_{n\to\infty} \frac{|\tau^n|}{|\pi^{k(n)}|} \geq 1$$

Proof. Let N be so large that for $n \geq N$ we have $\frac{|\pi^n|}{|\tau^n|} \geq 1$. If $\limsup_{n\to\infty} \frac{|\pi^n|}{|\tau^n|} < \infty$, it suffices to set $k(n) = n$. Otherwise, we pick our subsequence by the following rule:

$$k(n) = \inf\{k \geq n \,:\, |\pi^k| \leq |\tau^n|\}$$

Notice that for all n, $k(n) < \infty$ is guaranteed by the fact that $\lim_{n\to\infty} |\pi^n| = 0$. Then, it follows immediately from the definition of $k(n)$ that

$$\limsup_{n\to\infty} \frac{|\tau^n|}{\pi^{k(n)}} \geq 1$$

□

2.2 Quadratic and p-roughness

The key property that shall be useful in providing an intrinsic definition of p-th variation shall be p-roughness, the extension of quadratic roughness. The intuition behind p-roughness is that the cross p-order increments average to zero across coarse partitions. From now on, we shall make the assumption that $p \in \mathbb{Z}_{\geq 2}$.

Definition 2.3. (p-roughness) Let $\pi = (\pi^n)_{n=1}^{\infty}$ be a balanced sequence of partitions of $[0,T]$ with $|\pi^n| \to 0$. Let $0 < \beta < 1$. Suppose $X \in C[0,T]$ and has p-th variation along π. We say that X has the **p-roughness** property along π with coarsening index β on $[0,T]$ if for any β-coarsening A and any positive integers $k_1, \ldots, k_r$ with $k_1 + \ldots + k_r = p$ the following holds:

$$\sum_{j=1}^{N(A^n)} \sum_{\substack{p(n,j-1) \leq i_1,\ldots,i_r < p(n,j) \\ i_1,\ldots i_r \text{not all equal}}} \prod_{l=1}^{r} \left(X_{t_{i_l+1}} - X_{t_{i_l}}\right)^{k_l} \to 0 \tag{2.1}$$

as $n \to \infty$.

We denote the class of paths with p-roughness with coarsening index β along π by $R_\pi^{\beta,p}[0,T]$.

Notice that if $p = 2$, then Equation (2.1) becomes

$$\sum_{j=1}^{N(A^n)} \sum_{p(n,j-1)\leq i\neq k<p(n,j)} \left(X_{t_{i+1}} - X_{t_i}\right)\left(X_{t_{k+1}} - X_{t_k}\right) \to 0 \tag{2.2}$$

Which is precisely the definition of quadratic roughness of Cont and Das, showing that our proposed definition is a plausible generalisation of this property.

Testing whether a given trajectory satisfies Equation (2.1) is quite hard, so we provide a slightly simpler test below.

Theorem 2.1. (Quickly vanishing modulus of continuity test) Let $X \in C[0,T]$ with uniform modulus of continuity ω. Let π be a balanced sequence of partitions of $[0,T]$. If X admits p-variation along π and for any β-coarsening A of π we have

$$\sum_{j=1}^{N(A^n)} \sum_{\substack{p(n,j-1)\leq i_1,\ldots,i_r<p(n,j)\\ i_1,\ldots i_r \text{not all equal}}} \prod_{l=1}^{r} \omega\left(t_{i_l+1} - t_{i_l}\right) \to 0$$

Then $X \in R_\pi^{\beta,p}[0,T]$.

Proof. By the definition of the modulus of continuity of a continuous trajectory, we have:

$$\left|\sum_{j=1}^{N(A^n)} \sum_{\substack{p(n,j-1)\leq i_1,\ldots,i_r<p(n,j)\\ i_1,\ldots i_r \text{not all equal}}} \prod_{l=1}^{r} \left(X_{t_{i_l+1}} - X_{t_{i_l}}\right)^{k_l}\right| \leq \sum_{j=1}^{N(A^n)} \sum_{\substack{p(n,j-1)\leq i_1,\ldots,i_r<p(n,j)\\ i_1,\ldots i_r \text{not all equal}}} \prod_{l=1}^{r} \omega\left(t_{i_l+1} - t_{i_l}\right)$$
$$\to 0$$

□

That is, we can find functions with non-vanishing p-variation with prescribed modulus of continuity, and these functions will automatically satisfy the p-roughness property.

We remark that the conditions in the modulus of continuity test are not satisfied by Brownian motion, nor the Takagi-Landsberg class of functions defined in [59]. The conditions are, however, satisfied in the toy example we present below.

Example 2.1. Consider Takagi's blancmange function given by:

$$T(x) = \sum_{k=0}^{\infty} \frac{\phi(2^m x)}{2^m}$$

Where $\phi(y) = \text{dist}(y, \mathbb{Z})$. The modulus of continuity of T is [2]:

$$\omega(\delta) = C\delta \log\left(\frac{1}{\delta}\right)$$

For some $C > 0$. Let $\pi = (\pi^n)_{n=1}^{\infty}$ be a refining and balanced sequence of partitions with $A = (A^n)_{n=1}^{\infty}$ a β-coarsening of π for $0 < \beta < 1$. Then,

$$\begin{aligned}
R &= \sum_{j=1}^{N(A^n)} \sum_{p(n,j-1) \le i \ne j < p(n,j)} \omega(t_{i+1}^n - t_i^n)\omega(t_{j+1}^n - t_j^n) \\
&= C^2 \sum_{j=1}^{N(A^n)} \sum_{p(n,j-1) \le i \ne j < p(n,j)} (t_{i+1}^n - t_i^n) \log|t_{i+1}^n - t_i^n| (t_{j+1}^n - t_j^n) \log|t_{j+1}^n - t_j^n| \\
&\le C^2 \sum_{j=1}^{N(A^n)} \log^2 \underline{A_j^n} \sum_{p(n,j-1) \le i \ne j < p(n,j)} (t_{i+1}^n - t_i^n)(t_{j+1}^n - t_j^n) \\
&\le C^2 \sum_{j=1}^{N(A^n)} |A_j^n|^2 \log^2 \underline{A_j^n} \\
&\le C^2 |A^n| \log^2 \underline{A^n} \sum_{j=1}^{N(A^n)} |A_j^n| \\
&= C^2 |A^n| \log^2 \underline{A^n} \\
&\le C^2 c \underline{A^n} \log^2 \underline{A^n} \\
&\to 0
\end{aligned}$$

where we have used the fact that $\log^2 x$ is decreasing on $[0, 1]$ and that (A^n) is balanced. We conclude that T has quadratic roughness. The same argument can be repeated with p copies of the modulus of continuity to show that in fact T has p-roughness.

We conclude this section by exhibiting a large class of paths with 2-roughness.

Theorem 2.2. Let W be a standard Brownian motion defined on classical Wiener space. Let $\pi = (\pi^n)_{n=1}^{\infty}$ be a balanced and refining sequence of partitions satisfying, for $\nu > 2$

$$\lim_{n\to\infty} (\log n)^{\nu} |\pi^n| = 0$$

Then, for all $\beta \in (2/\nu, 1)$, the paths of W almost surely have the 2-roughness property along π with coarsening index β.

Proof. For a proof of this result, we refer the reader to Theorem 3.8 in [16]. □

The proof of the above result uses a combination of Lévy's modulus of continuity and the Hansen-Wright inequality to obtain sub-exponential concentration for the sums appearing in Equation (2.2). By additional careful estimates of the distances between a partition and its coarsened version, one can apply the Borel-Cantelli lemma to show that for Brownian motion, Equation (2.2) actually vanishes almost surely.

It is tempting to repeat the same procedure for fractional Brownian motion for $p = \frac{1}{H}$. However, obtaining the same sub-exponential concentration bounds, as was the case for Brownian motion, is not as easy. For instance, to use the Hansen-Wright inequality, one must use the independence of the increments in Brownian motion; for fractional Brownian motion one only has stationarity. Moreover, for $p > 2$, we have the product of more than two factors appearing in the expression for p-roughness. Intuitively, products are difficult to work with when seeking concentration of measure, as multiplication operations tend to involve a loss of "Lipschitzness" in the functionals we are trying to concentrate.

2.3 Uniqueness of p-th variation along balanced partitions

Lemma 2.2. Fix $0 < \beta < 1$ and let $\pi = (\pi^n)$ be a balanced sequence of partitions such that for $X \in C[0,T]$, the p-th variation $\langle X \rangle_{\pi}^{(p)}$ exists. If for every β-coarsening τ of π we have that $\langle X \rangle_{\pi}^{(p)} = \langle X \rangle_{\tau}^{(p)}$, then X has the p-roughness property.

Proof. Since τ coarsens π, we may construct the n-th level of τ through a sampling function p:

$$\tau^n = (0 = t^n_{p(n,0)} < t^n_{p(n,1)} < \ldots < t^n_{p(n,N(\tau^n))} = T)$$

Since X admits p-variation along τ, $\langle X\rangle_\pi^{(p)} = \langle X\rangle_\tau^{(p)}$, and τ coarsens π, for each $t \in [0,T]$ we have:

$$\begin{aligned}
0 &= \langle X\rangle_\pi^{(p)}(t) - \langle X\rangle_\tau^{(p)}(t) \\
&= \lim_{n\to\infty} \left(\sum_{\substack{s\in\pi^n \\ s\le t}} (X_{s'} - X_s)^p - \sum_{\substack{s\in\tau^n \\ s\le t}} (X_{s'} - X_s)^p \right) \\
&= \lim_{n\to\infty} \sum_{j=1}^{N(\tau^n)} \sum_{\substack{p(n,j-1)\le i_1,\dots,i_r<p(n,j) \\ i_1,\dots i_r \text{not all equal}}} \prod_{l=1}^{r} \left(X_{t_{i_l+1}} - X_{t_{i_l}} \right)^{k_l}
\end{aligned}$$

which is the definition of p-roughness, so that $X \in R_\pi^{\beta,p}[0,T]$. □

The theorem above shows that p-roughness is a necessary condition to obtain a unique p-variation along any coarsening of a balanced partition sequence.

Assumption. From now on, we assume that our balanced partitions have the further properties that: (i) $|\pi^n| \to 0$ and (ii) $\limsup_{n\to\infty} \frac{|\pi^n|}{|\pi^{n+1}|} < \infty$ or $\limsup_{n\to\infty} \frac{N(\pi^{n+1})}{N(\pi^n)} < \infty$. That is, we assume that the equivalence in the third condition in Proposition 2.1 holds.

Theorem 2.3. Let $\sigma = (\sigma^n)_{n=1}^\infty$ be a sequence of partitions satisfying the assumption above and, for some $0 \le \beta \le \alpha$, let $X \in C^\alpha[0,T] \cap R_\sigma^{\beta,p}[0,T]$ (i.e. X has p-roughness and is α-Hölder continuous). Then for any balanced sequence of partitions $\tau = (\tau^n)$ satisfying the assumption above, if X has p-th variation along τ then $\langle X\rangle_\tau^{(p)} = \langle X\rangle_\sigma^{(p)}$.

Proof. Throughout, we label points in the partitions σ^n using the letter s and those in the partition τ^n using the letter t. Following [16], we first consider the case $\limsup_n \frac{|\tau^n|}{|\sigma^n|} < 1$.

If we define $k(n) = \inf\{k \ge n \,:\, |\sigma^k| \le |\tau^n|\}$, then by Lemma 2.1, the sequences $(\sigma^{k(n)})_{n=1}^\infty$ and $(\tau^n)_{n=1}^\infty$ satisfy:

$$\limsup_{n\to\infty} \frac{|\tau^n|}{|\sigma^{k(n)}|} \ge 1 \qquad \limsup_{n\to\infty} \frac{|\tau^n|}{|\sigma^{k(n)-1}|} < 1$$

Thus, we may define:

$$l_n = \inf\left\{ l \ge n \,:\, N(\sigma^l) \ge N(\sigma^{k(n)})^{1/\beta} \right\}$$

so that $N(\sigma^{l_n-1}) \leq N(\sigma^{k(n)})^{1/\beta} \leq N(\sigma^{l_n})$. Since σ is balanced, we additionally get that (σ^{l_n}) is balanced, and so there are constants $c_1, c_2 > 0$ such that

$$c_1|\sigma^{l_n-1}| > |\sigma^{k(n)}|^{1/\beta} \geq c_2|\sigma^{l_n}|$$

For book-keeping, we define a function that encodes the ordering of τ^n with respect to σ^{l_n}; in particular, let:

$$p(n,k) = \inf\{m \geq 1 \,:\, s^{l^n}_m \in (t^n_k, t^n_{k+1}]\}$$

Then we have:

$$s^{l_n}_{p(n,k)-1} \leq t^n_k < s^{l_n}_{p(n,k)} < s^{l_n}_{p(n,k)+1} < \ldots < s^{l_n}_{p(n,k+1)-1} \leq t^n_{k+1} < s^{l_n}_{p(n,k+1)} \tag{2.3}$$

Our goal is now to show that $\langle X\rangle^{(p)}_{\tau^n}(T)$ and $\langle X\rangle^{(p)}_{\sigma^{l_n}}(T)$ have the same limits as $n \to \infty$. Decomposing $\Delta^n_k = X_{t^n_{k+1}} - X_{t^n_k}$ along the intermediate points σ^{l_n}, we get,

$$\underbrace{X_{t^n_{k+1}} - X_{t^n_k}}_{\Delta^n_k} = \underbrace{\left(X_{t^n_{k+1}} - X_{s^{l_n}_{p(n,k+1)-1}}\right)}_{D_k} - \underbrace{\left(X_{t^n_k} - X_{t^{l_n}_{p(n,k)-1}}\right)}_{B_k} + \underbrace{\sum_{i=p(n,k)}^{p(n,k+1)-1}\left[X_{s^{l_n}_i} - X_{s^{l_n}_{i-1}}\right]}_{C_k}$$

Then, using the labelling in (2.3), we get:

$$\begin{aligned}
\langle X\rangle^{(p)}_{\tau^n}(T) - \langle X\rangle^{(p)}_{\sigma^{l_n}}(T) &= \sum_{k=1}^{N(\tau^n)}\left[\Delta^p_k - \sum_{i=p(n,k)}^{p(n,k+1)-1}\left(X_{s^{l_n}_i} - X_{s^{l_n}_{i-1}}\right)^p\right] \\
&= \sum_{k=1}^{N(\tau^n)}\left[\Delta^p_k - C^p_k\right] + \sum_{k=1}^{N(\tau^n)}\left[C^p_k - \sum_{i=p(n,k)}^{p(n,k+1)-1}\left(X_{s^{l_n}_i} - X_{s^{l_n}_{i-1}}\right)^p\right]
\end{aligned}$$

Now, since p is an integer, we may factorise and bound the polynomial $\Delta_k^p - C_k^p$ to obtain:

$$\begin{aligned}
\left|\sum_{k=1}^{N(\tau^n)} [\Delta_k^p - C_k^p]\right| &\leq \sum_{k=1}^{N(\tau^n)} |\Delta_k - C_k||\Delta_k^{p-1} + \Delta_k^{p-2}C_k + \ldots + C_k^{p-1}| \\
&\leq \sum_{k=1}^{N(\tau^n)} pM^{p-1}|\Delta_k - C_k| \qquad (M = 2\left\|X\right\|_\infty) \\
&= \sum_{k=1}^{N(\tau^n)} pM^{p-1}|D_k - B_k| \\
&\leq \sum_{k=1}^{N(\tau^n)} pM^{p-1}\left(|D_k| + |B_k|\right) \\
&\leq \sum_{k=1}^{N(\tau^n)} 2pM^{p-1}\left\|X\right\|_\alpha |\sigma^{l_n}|^\alpha \\
&\leq 2pN(\tau^n)M^{p-1}\left\|X\right\|_\alpha |\sigma^{l_n}|^\alpha \\
&\leq 2cN(\sigma^{k(n)})pM^{p-1}\left\|X\right\|_\alpha |\sigma^{k(n)}|^{\alpha/\beta} \\
&= KN(\sigma^{k(n)})|\sigma^{k(n)}|^{\alpha/\beta} \\
&\to 0
\end{aligned}$$

since $\alpha \geq \beta$, K is a constant independent of n, and $(\sigma^{k(n)})$ is a balanced partition with vanishing mesh. Note that $M = 2\left\|X\right\|_\infty$ comes from the triangle inequality applied to the increments Δ_k. Furthermore, telescoping the term C_k and expanding it out, we obtain:

$$\begin{aligned}
&\sum_{k=1}^{N(\tau^n)} \left[C_k^p - \sum_{i=p(n,k)}^{p(n,k+1)-1} \left(x(s_i^{l_n}) - x(s_{i-1}^{l_n})\right)^p\right] \\
&= \sum_{k=1}^{N(\tau^n)} \left[\left(X_{s^{l_n}_{p(n,k+1)-1}} - X_{s^{l_n}_{p(n,k)-1}}\right)^p - \sum_{i=p(n,k)}^{p(n,k+1)-1} \left(X_{s_i^{l_n}} - X_{s_{i-1}^{l_n}}\right)^p\right] \\
&= \sum_{k=1}^{N(\tau^n)} \sum_{\substack{i_1,\ldots,i_r = p(n,k) \\ i_t \text{ not all equal}}}^{p(n,k+1)-1} \prod_{t=1}^{r} \left(X_{s^{l_n}_{i_t}} - X_{s^{l_n}_{i_t-1}}\right)^{k_t}
\end{aligned}$$

To send this quantity to zero, we must now show that the sequence of partitions given by $A^n = (0 = s_0^{l_n} < s_{p(n,1)-1}^{l_n} < s_{p(n,2)-2}^{l_n} < \ldots < s_{p(n,N(\tau^n))-1}^{l_n} = T)$ is a β-coarsening of

(σ^{l_n}). First, for large n, we have $\frac{1}{2}\underline{\tau^n} \leq \underline{A_n} \leq |A^n| \leq 2|\tau^n|$, so that $(A^n)_{n=1}^{\infty}$ is balanced. Moreover, there exist constants $k, k', k'' > 0$ such that

$$\begin{aligned}
0 &< k \liminf_{n\to\infty} \frac{|\tau^n|}{\sigma^{k(n)}} \\
&\leq \liminf_{n\to\infty} \frac{|\tau^n|}{|\sigma^{l_n}|^\beta} \\
&\leq \limsup_{n\to\infty} \frac{|\tau^n|}{|\sigma^{l_n}|^\beta} \\
&\leq k' \limsup_{n\to\infty} \frac{|\tau^n|}{|\sigma^{k(n)+1}|} \\
&\leq k'' \limsup_{n\to\infty} \frac{|\tau^n|}{|\sigma^{k(n)}|} \times \frac{|\sigma^{k(n)}|}{|\sigma^{k(n)+1}|} \\
&< \infty
\end{aligned}$$

Thus, (A^n) is a β-coarsening of (σ^{l_n}). The p-roughness property of X along (σ^{l_n}) then implies that:

$$\sum_{k=1}^{N(\tau^n)} \sum_{\substack{i_1,\ldots,i_r = p(n,k) \\ i_t \text{ not all equal}}}^{p(n,k+1)-1} \prod_{t=1}^{r} \left(X_{s_{i_t}^{l_n}} - X_{s_{i_t-1}^{l_n}} \right)^{k_t} \to 0 \qquad (\text{as } n \to \infty)$$

Hence,

$$\langle X \rangle_\sigma^{(p)}(T) - \langle X \rangle_\tau^{(p)}(T) = \lim_{n\to\infty} \langle X \rangle_{\sigma^n}^{(p)}(T) - \langle X \rangle_{\tau^n}^{(p)}(T) = 0$$

The result for $t < T$ holds similarly, with the addition of the rightmost term in each partition, which converges to zero as well, so that in fact we have $\langle X \rangle_\sigma^{(p)}(t) = \langle X \rangle_\tau^{(p)}(t)$ for all $t \in [0, T]$.

Finally, if $\limsup_n \frac{|\tau^n|}{|\sigma^n|} \geq 1$, then we may let $r(n) = \inf\{r \geq n : |\tau^r| < |\sigma^n|\}$, which is finite as $|\tau^n| \to 0$. Set $\pi^n = \tau^{r(n)}$, so that that $\limsup_{n\to\infty} \frac{|\pi^n|}{|\sigma^n|} < 1$. By the result we just proved, we get that $\langle X \rangle_\sigma^{(p)}(t) = \langle X \rangle_\pi^{(p)}(t) = \langle X \rangle_\tau^{(p)}(t)$ for all $t \in [0, T]$, where the last equality holds because π is a subsequence of τ.

□

We now investigate an example for which p-roughness fails. We remark that testing the definition of p-roughness directly is difficult and an easier approach is to test when the conclusion in the previous theorem fails to hold.

Example 2.2. We show that a certain generalised Takagi function fails to have quadratic roughness. Consider the Faber-Schauder basis for $C[0,1]$ given by:

$$e_{0,0}(t) = (\min(t, 1-t))^+$$
$$e_{m,k}(t) = 2^{-m/2} e_{0,0}(2^m t - k)$$

We define the following two expansions in terms of these bases:

$$X = \sum_{m=0}^{\infty} \sum_{k=0}^{2^m-1} e_{m,k}$$
$$Y = \sum_{m=0}^{\infty} \sum_{k=0}^{2^m-1} (-1)^m e_{m,k}$$

By the Weierstrass M-test, these two series are absolutely and uniformly convergent, so that $X, Y \in C[0,1]$. It was proven in Proposition 2.7 of [69] that if $\pi = (\pi^n)_{n=1}^{\infty}$ is the dyadic partition of $[0,1]$, $\tau = (\pi^{2n})_{n=1}^{\infty}$, and $\sigma = (\pi^{2n+1})_{n=1}^{\infty}$ then:

$$\langle X + Y \rangle_{\tau}^{(2)}(t) = \frac{4}{3}t$$
$$\langle X + Y \rangle_{\sigma}^{(2)}(t) = \frac{8}{3}t$$

Notice that the 8-adic partition of $[0,1]$, say $A = (A^n)_{n=1}^{\infty}$ is a (balanced) subsequence of τ, so that the quadratic variation of $X + Y$ along A is also $\frac{4}{3}t$. Furthermore, note that τ and σ are $\frac{1}{2}$-coarsenings of A.

Since $X + Y$ has different quadratic variations along coarsenings of the balanced partition A, it follows that $X + Y$ does not satisfy the quadratic roughness property.

Once one has access to a class of functions with prescribed quadratic variation, it is often desirable to expand this class by performing a suitable time-change, as one does in classical stochastic calculus. It was observed in [58] that this procedure can be problematic, as a time-changed function will admit quadratic variation under a time changed sequence of partitions, and even when it does, it is uncertain whether control of this quadratic variation can be attained. As an application of p-roughness, we can obtain control of p-variation under smooth time reparametrisations.

Theorem 2.4. Let $\pi = (\pi^n)_{n=1}^{\infty}$ be a balanced sequence of partition, $g \in C^1[0,T]$ with $\inf g' > 0$, and $0 < \beta < 1$. If $X \in R_\pi^{\beta,p}[0, g(T)]$, then $X \circ g \in V_p(\pi)$ and:

$$\langle X \rangle_\pi^{(p)}(g(t)) = \langle X \circ g \rangle_\pi^{(p)}(t)$$

Proof. We first note that $g(\pi) = (g(\pi^n))_{n=1}^{\infty}$ is a balanced sequence of partitions for $[g(0), g(T)]$. Indeed, if we write $\pi^n = (0 = t_1^n < t_2^n < \ldots < t_{N(\pi^n)}^n = T)$, by the mean-value theorem we can find numbers $u_k^n, v_k^n \in (t_k^n, t_{k+1}^n)$ such that:

$$\frac{|g(\pi^n)|}{\underline{g(\pi^n)}} = \frac{\sup_{t_k \in \pi^n} g'(u_k^n)(t_{k+1}^n - t_k^n)}{\inf_{t_k \in \pi^n} g'(v_k^n)(t_{k+1}^n - t_k^n)}$$

Taking the limit superior, we get:

$$\begin{aligned}
\limsup_{n\to\infty} \frac{|g(\pi^n)|}{\underline{g(\pi^n)}} &\leq \left| \limsup_{n\to\infty} \frac{\sup_{\pi^n} g'(u_k^n)}{\inf_{\pi^n} g'(v_k^n)} \right| \left| \limsup_{n\to\infty} \frac{\sup_{t_k \in \pi^n}(t_{k+1}^n - t_k^n)}{\inf_{t_k \in \pi^n}(t_{k+1}^n - t_k^n)} \right| \\
&\lesssim \frac{\sup_{t\in[0,T]} g'(t)}{\inf_{t\in[0,T]} g'(t)} \\
&< \infty
\end{aligned}$$

Which means that $g(\pi)$ is balanced.

Arguing again by the mean value theorem, for each n, k get the $\xi_k^n \in (t_k^n, t_{k+1}^n)$ that achieves the average slope for g, and estimate:

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\sup_{\pi^n}(g(t_{k+1}^n) - g(t_k^n))}{\sup_{\pi^{n+1}}(g(t_{k+1}^{n+1}) - g(t_k^{n+1}))} &= \limsup_{n\to\infty} \frac{\sup_{\pi^n} g'(\xi_k^n)(t_{k+1}^n - t_k^n)}{\sup_{\pi^{n+1}} g'(\xi_k^{n+1})(t_{k+1}^{n+1} - t_k^{n+1})} \\
&\leq \frac{\sup_{t\in[0,T]} g'(t)}{\inf_{t\in[0,T]} g'(t)} \limsup_{n\to\infty} \frac{|\pi^n|}{|\pi^{n+1}|} \\
&< \infty
\end{aligned}$$

A final application of the mean value theorem shows that $g(\pi)$ and π are asymptotically comparable. Then, the assumptions in Theorem 2.3 are satisfied for $X \circ g$, which implies that

$$\langle X \rangle_\pi^{(p)}(g(t)) = \langle X \circ g \rangle_\pi^{(p)}(t)$$

□

We end this section by putting all the ingredients above together, into a partition-free definition of p-variation.

Proposition 2.2. There exists a set $K[0,T] \subset C[0,T]$ and a map

$$\langle \cdot \rangle^{(p)} : K \to C[0,T]$$

Such that for all balanced partitions π of $[0,T]$ we have $\langle X \rangle_{\pi}^{(p)} = \langle X \rangle^{(p)}$. We call $\langle X \rangle^{(p)}$ the (intrinsic) p-variation of the trajectory X.

Proof. In light of Theorem 2.3, it suffices to set $K[0,T]$ to be the set of $\frac{1}{2} - \epsilon$-Hölder continuous functions which satisfy the p-roughness property along balanced sequences of partitions. Finally, to construct $\langle X \rangle^{(p)}$, it suffices to pick the quadratic variation of X along any balanced partition sequence of $[0,T]$. □

2.4 Applications to pathwise Itô calculus

In this section we show that the p-roughness property is crucial in obtaining an intrinsic definition of the pathwise integral and the pathwise local time.

2.4.1 Uniqueness of pathwise integral

Given that we now have access to an intrinsic map for p-variation, we may proceed to construct an intrinsic rough pathwise Itô integral. To obtain such intrinsic map, it will be profitable to have a computational formula for the p-variation of a rough pathwise integral. First, we show that p-variation is undisturbed by perturbations of bounded variation.

Proposition 2.3. Fix a refining sequence of partitions $\pi = (\pi_n)_{n=1}^{\infty}$. Let $X \in C[0,T]$ and let $A \in C[0,T] \cap BV[0,T]$. Then,

$$\langle X + A \rangle_{\pi}^{(p)}(t) = \langle X \rangle_{\pi}^{(p)}(t) \qquad t \in [0,T]$$

Proof. For each $t \in [0,T]$,

$$\begin{aligned}\left|\langle X+A\rangle_\pi^{(p)}(t) - \langle X\rangle_\pi^{(p)}(t)\right| &= \left|\lim_{n\to\infty}\sum_{\substack{s\in\pi^n\\ s\le t}}(X_{s'}-X_s+A_{s'}-A_s)^p - \langle X\rangle_\pi^{(p)}(t)\right| \\ &= \left|\lim_{n\to\infty}\sum_{\substack{s\in\pi^n\\ s\le t}}\left[(X_{s'}-X_s)^p + \sum_{k=1}^{p}(X_{s'}-X_s)^{p-k}(A_{s'}-A_s)^k\right] - \langle X\rangle_\pi^{(p)}(t)\right|\end{aligned}$$

Notice that for each fixed $1 \le k \le p$

$$\begin{aligned}\sum_{\substack{s\in\pi^n\\ s\le t}}\binom{p}{k}(X_{s'}-X_s)^{p-k}(A_{s'}-A_s)^k &\le p!\sup_{s\in\pi^n}|(X_{s'}-X_s)^{p-k}(A_{s'}-A_s)|^{k-1}\sum_{s\in\pi^n}|(A_{s'}-A_s)| \\ &\le p!\sup_{s\in\pi^n}|(X_{s'}-X_s)^{p-k}(A_{s'}-A_s)|^{k-1}\|A_t\|_{var} \\ &\to 0\end{aligned}$$

where $\|A_t\|_{var} < \infty$ is the total variation of A_t, and the limit goes to zero by the uniform continuity of X_t and A_t on $[0,T]$. Thus, the limits in the approximation above can be taken term-by-term. Since, X admits p-th variation along π, we get for each $t \in [0,T]$ that

$$\lim_{n\to\infty}\sum_{\substack{s\in\pi^n\\ s\le t}}(X_{s'}-X_s)^p - \langle X\rangle_\pi^{(p)}(t) = 0$$

so that finally, after taking limits, $\left|\langle X+A\rangle_\pi^{(p)}(t) - \langle X\rangle_\pi^{(p)}(t)\right| = 0$, or—more compactly—$\langle X+A\rangle_\pi^{(p)} = \langle X\rangle_\pi^{(p)}$ □

Now we have all the ingredients necessary to find an intrinsic definition of the rough Itô integral.

Theorem 2.5. (Uniqueness of the Itô-Föllmer-Cont-Perkowski integral) There exists a unique map:

$$I : C^p(\mathbb{R}) \times K[0,T] \to K[0,T]$$

Given by:

$$I(f,X)_t = \int_0^t f'(X_s)dX_s,$$

such that for all balanced partition sequences $\pi \in \mathbb{B}[0,T]$, the following holds:

$$I(f,X)_t = \int_0^t f'(X_s)dX_s^\pi = \lim_{n\to\infty} \sum_{\substack{s\in\pi^n \\ s\leq t}} \sum_{k=1}^{p-1} \frac{f^{(k)}}{k!}(X_{s'} - X_s)^k \tag{2.4}$$

Moreover, the p-th variation of $I(f,X)$ along any balanced partition π is given by:

$$\langle I(f,X)\rangle_\pi^{(p)} = \left\langle \int_0^t f'(X_s)dX_s \right\rangle_\pi^{(p)} = \int_0^t f'(X_s)^p d\langle X\rangle_\pi^{(p)}(s) \tag{2.5}$$

Proof. Let π be a balanced sequence of partitions for $[0,T]$, and let σ be another balanced sequence of partitions. Thus, by Theorem 2.2, any trajectory $X \in K[0,T]$ admits the same quadratic variation along π and σ. Then, we may use two applications of Theorem 1.1 to get:

$$\begin{aligned}
\int_0^t f'(X_s)dX_s^\pi &= f(X_t) - f(X_0) - \frac{1}{p!}\int_0^t f^{(p)}(X_s)d\langle X\rangle_\pi^{(p)}(s) \\
&= f(X_t) - f(X_0) - \frac{1}{p!}\int_0^t f^{(p)}(X_s)d\langle X\rangle_\sigma^{(p)}(s) \\
&= \int_0^t f'(X_s)dX_s^\sigma
\end{aligned}$$

Which implies that on $K[0,T]$ we may define a unique, or intrinsic, pathwise Itô integral.

To show that 2.5 holds, we apply Taylor's theorem in a non-anticipating manner:

$$\begin{aligned}
\left\langle \int_0^t f'(X_s)dX_s \right\rangle_\pi^{(p)} &= \langle f(X_t) + A_t \rangle_\pi^{(p)} && A_t \in BV[0,T] \\
&= \langle f(X_t) \rangle_\pi^{(p)} \\
&= \lim_{n\to\infty} \sum_{\substack{s\in\pi^n \\ s\le t}} [f(X_{s'}) - f(X_s)]^p \\
&= \lim_{n\to\infty} \sum_{\substack{s\in\pi^n \\ s\le t}} f'(X_s)^p (X_{s'} - X_s)^p + \underbrace{O((X_{s'} - X_s)^q)}_{\to 0} && (q > p) \\
&= \lim_{n\to\infty} \sum_{\substack{s\in\pi^n \\ s\le t}} f'(X_s)^p (X_{s'} - X_s)^p \\
&= \int_0^t f'(X_s)^p d\langle X \rangle_\pi^{(p)}(s)
\end{aligned}$$

□

In Föllmer's calculus, one may use a simpler argument, relying only on the mean value theorem to obtain

$$\left\langle \int_0^t f'(X_s)dX_s \right\rangle_t^{(2)} = \int_0^t f'(X_s)^2 d\langle X \rangle_s^{(2)}$$

This formula is remarkably useful in robust finance; for instance, it plays a key role when applied alongside the Breeden-Litzenberger formula to find a replicating portfolio to value log contracts on stocks. These log contracts then form the basis for the VIX index. If more exotic contracts are liquidly traded in the markets which are built on the basis of rough trajectories, then our formulas above can be used to price said contracts.

2.4.2 Uniqueness of local time

In several applications in stochastic calculus, one is interested in applying Itô's formula to the distance of a Brownian motion to the origin, $||B_t||$. For a one-dimensional Brownian motion, this amounts to applying Itô's formula to the function $f(x) = |x|$, which, of course,

cannot be done as f is not differentiable. All is not lost. By approximating the function f by C^2 functions, we may in fact show that the following equation holds:

$$|B_t - a| = |a| + \int_0^t \text{sgn}(B_s - a)dB_s + \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t \chi_{(a-\epsilon, a+\epsilon)}(B(s))ds \tag{2.6}$$

where $a \in \mathbb{R}$. The last term in the right-hand side has its own name.

Definition 2.4. (Brownian local time) The random variable defined by

$$L_a(t) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t \chi_{(a-\epsilon, a+\epsilon)}(B(s))ds \tag{2.7}$$

is called the local time of a Brownian motion at a, up to and including t.

Intuitively, fixing $\omega \in \Omega$, we can interpret $L_a(t)(\omega)$ as the amount of time the path $B_t(\omega)$ spends at a up to time t. With the newly introduced terminology, we may phrase Equation (2.6) in its more usual form.

Theorem 2.6. (Tanaka's formula) Let B_t be a standard Brownian motion on $\mathbb{R}$. Then, for any $a \in \mathbb{R}$

$$|B_t - a| = |a| + \int_0^t \text{sgn}(B_s - a)dB_s + L_a(t) \tag{2.8}$$

where $L_a(t)$ is the Brownian local time at a.

Unsurprisingly, there are semi-martingale and pathwise versions of local time. In her unpublished Master's thesis [75], Wuermli proved a pathwise version of Tanaka's formula, where one loosens the requirement that $f \in C^2$ to requiring that f admits second-order weak derivatives in L^2.

Cont and Perkowski (and previously Davis et al, [21]) also consider extensions of their change-of-variable formula for less regular functions. To that end, we consider the following definition introduced in [18]:

Definition 2.5. (Local time of order p) Let $p \in 2\mathbb{N}$ and $q \in [1, \infty]$. A trajectory $X \in C[0,T]$ admits L^q-local time of order $p-1$ along a refining sequence of partitions $\pi = (\pi^n)_{n=1}^{\infty}$, where $\pi^n = (0 = t_0^n < t_1^n < \ldots < t_{N(\pi^n)} = T)$ if the sequence of maps

$$L_t^{\pi_n, p-1}(u) = \sum_{t_j^n \in \pi^n} 1_{(X_{t_j^n}, X_{t_{j+1}^n}]}(u) |X_{t_{j+1}^n} - u|^{p-1} \tag{2.9}$$

converges weakly in $L^q(\mathbb{R})$ to a weakly continuous map $L : [0,T] \to L^q(\mathbb{R})$. Note that the intervals in the indicator function of (2.9) are sorted to be non-degenerate. The set of continuous paths admitting this property is denoted by $\mathcal{L}_p^q(\pi)$.

The function L intuitively measures how much p-variation the path X accumulates at the level u. As an extension to Wuermli's result, the following local time formula can be obtained for L:

Theorem 2.7. (Pathwise Tanaka formula) Let $p \in 2\mathbb{N}$ and $q \in [1,\infty]$. Let q^* be the Hölder conjugate of q, so that $(L^q)^* \cong L^{q^*}$. Let $f \in C^{p-1}(\mathbb{R})$ and let $f^{(p-1)}$ admit a weak derivative $f^{(p)}$ in L^{q^*}. Then, for any $X \in \mathcal{L}_p^q(\pi)$ we have:

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)dX_s + \frac{1}{(p-1)!}\int_{\mathbb{R}} f^{(p)}(x)L_t(x)d\mu \tag{2.10}$$

Where μ is the Lebesgue measure.

As in the previous section, we may now apply our property of p-roughness to find an intrinsic version of this local time theorem.

Theorem 2.8. Let π be a balanced and refining (i.e. $\pi^n \subset \pi^{n+1}$) sequence of partitions. For $X \in K[0,T] \cap \mathcal{L}_p^q(\pi)$ the pathwise Tanaka formula remains invariant across balanced partitions. Moreover, the local time function is itself identical across all balanced sequences of partitions.

Proof. Let L_t and L'_t be the local times of order p obtained by taking the limit along two partitions π and π', respectively, as in the assumptions of this theorem. By applying the pathwise Tanaka formula twice (Theorem 2.7) and the uniqueness of the rough integral (Theorem 2.5), we get:

$$\begin{aligned}\frac{1}{(p-1)!}\int_{\mathbb{R}} f^{(p)}(x)L_t(x)d\mu &= f(X_t) - f(X_0) - \int_0^t f'(X_s)dX_s^{\pi} \\ &= f(X_t) - f(X_0) - \int_0^t f'(X_s)dX_s^{\pi'} \\ &= \frac{1}{(p-1)!}\int_{\mathbb{R}} f^{(p)}(x)L'_t(x)d\mu\end{aligned}$$

Thus, for each $t \in \mathbb{R}^+$

$$\int_{\mathbb{R}} f^{(p)}(x)L_t(x)d\mu = \int_{\mathbb{R}} f^{(p)}(x)L'_t(x)d\mu \tag{2.11}$$

This shows that the local time term in extending Itô's formula to low regularity functionals remains invariant across balanced partitions.

More is true: in fact we show that $L_t = L'_t$ for almost every t. By Remark 2.10 in [51], we have the following occupation time formula for any continuous function g:

$$p \int_0^t g(x) L_t(x) dx = \int_0^t g(X_s) d \langle X \rangle_s^{(p)}$$

This implies that $\int_0^t g(x) L_t(x) dx = \int_0^t g(x) L'_t(x) dx$ for any continuous g; but by the fundamental theorem of calculus (or Lebesgue's differentiation theorem), that implies that $L_t = L'_t$ almost everywhere, as desired. □

We conclude this chapter with a few departing open questions:

1. Does p-roughness hold for fractional Brownian motion with Hurst exponent $H = 1/p$?
2. Can we find a large family of continuous functions in a probability-free manner which has prescribed quadratic variation and admits p-roughness?
3. Can the notion of p-roughness be extended to multivariate trajectories?
4. Can p-roughness be extended to non-integer exponents?

Chapter 3

Deep learning the roughness of trajectories under $\mathbb{P}$ and $\mathbb{Q}$

The objective of this chapter is to investigate whether physical asset trajectories exhibit elements of roughness and, if so, whether we can directly measure how rough these trajectories are. We will also be interested in measuring these characteristics in the risk-neutral world. To distinguish the physical and risk-neutral realms, we introduce the notation $\mathbb{P}$ and $\mathbb{Q}$ to denote the statistical and risk-neutral measures, respectively. To motivate statistical roughness, we recall some properties of fractional Brownian motion.

3.1 A rough model: fractional Brownian motion

Standard Brownian motion is a process with continuous paths with independent Gaussian increments. To generalise this process, the property of independent increments can be relaxed.

Definition 3.1. Let $H \in (0,1)$. A fractional Brownian motion with Hurst parameter H is a centred Gaussian process $B^{(H)} = \{B_t^{(H)}\}_{t\geq 0}$ with the covariance structure:

$$\mathbb{E}\left(B_t^{(H)} B_s^{(H)}\right) = \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H}) \tag{3.1}$$

Observe that for $H = \frac{1}{2}$ we recover standard Brownian motion. Moreover, applying Kolmogorov's continuity criterion and the Garsia-Rodemic-Rumsey lemma, one can discover that the paths of $B^{(H)}$ are almost surely $H - \epsilon$ Hölder continuous, for all $\epsilon > 0$ [63].

Thus, the Hurst parameter controls the regularity of fractional Brownian motion. Because of that, we shall say that fractional Brownian motion with $H < \frac{1}{2}$ is "rougher" than standard Brownian motion and when $H > \frac{1}{2}$ we informally say it is "smoother" (see Figure 3.1). Elsewhere in the literature or in trading circles the terminology "anti-persistent" and "trending" may be found for rough and smooth time series, respectively.

For completeness, we list down other elementary properties of fractional Brownian motion; we refer the reader to [8, 62, 63] for more detailed treatises of fractional Brownian motion.

Proposition 3.1. Let $B^{(H)}$ be a fractional Brownian motion with Hurst parameter $H \in (0, 1)$. Then:

1. $B^{(H)}$ has stationary increments satisfying $\mathbb{E}\left(|B^{(H)}_{t+\Delta} - B^{(H)}_t|^q\right) = K_q \Delta^{qH}$ for any $t \in \mathbb{R}, \Delta \geq 0$, and $q > 0$;
2. $B^{(H)}$ is self-similar: $\mathcal{L}(B^H_{at}) = \mathcal{L}(a^H B^H_t)$;
3. $B^{(H)}$ is a semi-martingale if and only if $H = \frac{1}{2}$.

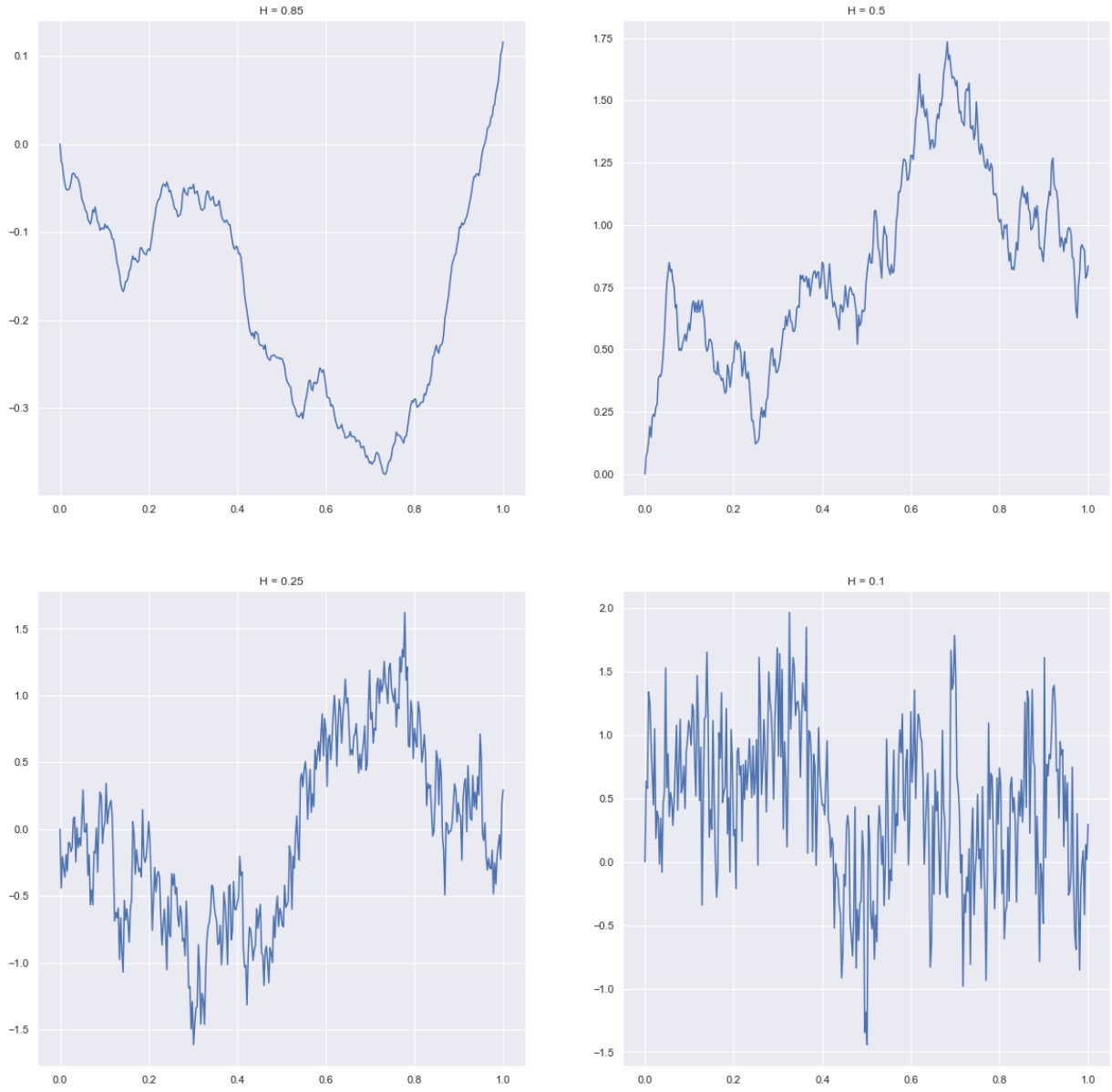

Figure 3.1: Fractional brownian motion with different Hurst exponents

Oftentimes, it is useful to represent a fractional Brownian motion in terms of standard Brownian motion. Several representations exist, but perhaps one of the most useful ones is the Mandelbrot-van Ness representation [57]:

Proposition 3.2. (Mandelbrot-van Ness) A fractional Brownian motion with Hurst parameter H may be represented as the average of a standard Brownian motion against a power-law kernel:

$$B_t^{(H)} = C_H \left(\int_0^t (t-s)^{H-\frac{1}{2}} dB_s + \int_{-\infty}^0 \left[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right] dB_s \right) \tag{3.2}$$

Where $C_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}$.

We remark that more modern treatments regard the above integral as an integral against white noise, the distributional derivative of Brownian motion. A related process is called the Holmgren-Riemann-Liouville fractional integral:

$$B_t^{(H)} = \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} dB_s \tag{3.3}$$

This integral will appear when describing some $\mathbb{Q}$-models for asset prices.

But not all data arises from fractional Brownian motion. So we wish to interpret the parameter H outside of the realm of Gaussian processes. So what interpretation do we give it? From the discussion above it is clear that if a trader has access to continuously sampled data, then—in the $\mathbb{P}$ world—they would observe fractal or self-similarity properties, which directly circle back to H, which governs the Hölder regularity of a trajectory.

Without access to continuously sampled data, we may need proxies for H. One such proxy arises from options market data. In particular, if H is an abstract parameter that governs a process's roughness or irregularity, then this should be invariant under change of measure. Indeed, for Brownian motion, change of measure to an equivalent measure leaves the Hölder regularity unchanged, because the set $C^{H^-}[0,T]$ occurs with probability one. Thus, if option prices are consistent with the prices suggested by a model containing a roughness parameter, then we may posit that our trader actually observed discretely sampled points from a physical trajectory which exhibits roughness consistent to that of option prices.

Thus, to posit rough physical models, we actually need to revert to the artificial risk-neutral world first.

3.2 (Rough) Stochatic volatility

Here we present two models of interest which recently appeared in the stochastic volatility literature.

Rough Bergomi model

Under the risk-neutral measure, the rough Bergomi model is given by:

$$dS_t = -\frac{1}{2}v_t dt + \sqrt{v_t}dW_t \tag{3.4}$$

$$v_t = \xi_0(t)\mathcal{E}\left(\sqrt{2H}\nu \int_0^t (t-s)^{H-1/2} dB_s\right) \tag{3.5}$$

$$\langle W, B\rangle_t = \rho t \tag{3.6}$$

where $\mathcal{E}$ is the Dóleans-Dade exponential. As usual, we regard (S, v) as the log-stock-variance pair; $\xi_u(t) = \mathbb{E}\left(v_t \,|\, \mathcal{F}_u\right)$ the variance curve for $t \geq u$, ν a vol-of-vol parameter, $\rho \in [-1, 1]$ is stock-volatility correlation, and $H \in \left(0, \frac{1}{2}\right)$ the Hurst index of volatility.

In [6], the authors perform several formal manipulations of the Mandelbrot-van Ness representation of fractional Brownian motion and the variance curve representation of $\sigma_t = \sigma \exp(X_t)$, where X_t is a fractional Ornstein-Uhlenbeck process, to arrive at the rough Bergomi model.

This model is inspired by the variance-curve model proposed by Lorenzo Bergomi in his book (Chapter 7, [7]). To motivate it we first let $\sigma_{BS}(k, \tau)$ denote the Black-Scholes call option implied volatiliy at log-moneyness k and time-to-maturity τ. The implied volatility is the required volatility in the Black-Scholes model to produce market prices. Bergomi notes that the volatility skew in a variance curve model

$$\Psi(\tau) = \left|\frac{\partial}{\partial k}\sigma_{BS}(k, \tau)\right|_{k=0} \qquad k = \log\frac{K}{S_t}, \qquad \tau = T - t \tag{3.7}$$

is strongly dependent on the choice of moving-average kernel in the volatility equation. That is, if one has a volatility model containing the term $\mathcal{E}\left(\int_0^t \phi(t, s)dB_s\right)$, then the volatility skew will depend strongly in the choice of ϕ. Bergomi's original model uses a decaying exponential kernel, which in turn induces a superposition of decaying exponential curves for the volatility skew. Inspired by this observation and by the observation that the market implied volatility skew follows a power law, one may instead use a power law kernel in the specification of the model, as in Equation (3.5). Indeed, this is what the rough Bergomi model proposes, as the argument of the Wick exponential is itself a power law kernel applied to a Brownian motion.

Heston model

Under $\mathbb{Q}$, the Heston model is given by:

$$dS_t = S_t\sqrt{v_t}dW_t \tag{3.8}$$
$$dv_t = \kappa(\theta - v_t)dt + \nu\sqrt{v_t}dB_t \tag{3.9}$$
$$\langle W, B\rangle_t = \rho t \tag{3.10}$$

The reason for the popularity of the Heston model is that its characteristic function can be computed easily [43]. With access to its characteristic formula, one can perform Fourier pricing to recover the prices of European calls and puts on an asset that behaves under the Heston dynamics [13]. A rough version of the Heston model exists; finding computational methods to simulate its paths and numerically evaluate its characteristic function are active areas of research [25].

While the Heston model has realistic dynamics for the stock and volatility process, it does not replicate the features of observed volatility surfaces as well as desired [35].

3.3 Deep learning set-up

In the past 20 years, neural networks have gained prominence for their ability to solve several high-dimensional tasks, including image recognition and language analysis. Their ability to tackle these high-dimensional statistical learning problems hinges on a certain density property of the class of neural networks. First, we define a neural network.

Definition 3.2. Let $L \in \mathbb{N}$ and $(N_1, \ldots, N_L) \in \mathbb{N}^L$ be fixed parameters. Let $W^i : \mathbb{R}^{N_i} \to \mathbb{R}^{N_{i+1}}$ be affine maps for $1 \leq i \leq L-1$. Set $F_i = s_i \circ W^i$, where $s_i : \mathbb{R} \to \mathbb{R}$ is an activation function applied component-wise. Then a **neural network** is a function $F_W : \mathbb{R}^{N_1} \to \mathbb{R}^{N_L}$ given by the composition:

$$F_W = F_L \circ \ldots \circ F_1 \tag{3.11}$$

In the definition above, we refer to L as the depth of the network, and the tuple $(N_1, \ldots, N_L)$ as the neurons in the network. The layers F_1 and F_L denote the input and output layers, respectively. The layers in between, $F_2, \ldots, F_{L-1}$, are called hidden layers. For example, we may specialise the definition above to the case where we have a single

hidden layer, a single activation function, d input variables, and a single output variable (as may be the case in a regression problem), so that our neural network has the form:

$$F_W(x) = \sum_{j=1}^{m} \alpha_j s(y_j^T x + \theta_j) \qquad y_j \in \mathbb{R}^d, \theta_j, \alpha_j \in \mathbb{R} \tag{3.12}$$

If we denote the vector space of all neural networks of the form in Equation (3.12) by $\mathcal{N}_s$, then Cybenko's universal approximation theorem [20] gives mild conditions under which $\mathcal{N}_s$ is uniformly dense in the space of continuous functions. Such universal approximation theorems for different classes of networks are an active area of study: if one can prove a universal approximation by a class of networks to a given problem, then one has reason to believe that neural networks are a good way to solve the problem at hand.

Bounds on the errors of neural networks with different architectures have been studied since their advent (see [5, 26] and the references therein). Informally, the conclusion these studies often reach is that deep networks have higher expressibility than shallow networks, although there are diminishing marginal returns after the addition of the first few layers. We use these findings to inform the architecture of our networks in later sections.

3.4 Deep learning H under the risk-neutral measure

In this section we explore a deep learning (or rather, deep calibration) framework under which we can determine the value of H. In financial model calibration, the analyst prescribes a market model parametrised by a vector θ; after observing prices for liquidly traded securities, calibration refers to finding the choice of θ which makes the observed market prices most plausible. In this section, we focus on learning H under the risk-neutral measure, which means that our main source of data will be options data.

3.4.1 Prevailing learning and calibration framework

A recent well-received method of calibration is the grid-based method introduced by Horvath, Muguruza, and Tomas in [45]. In their paper, the authors provide a general framework to learn H and calibrate a stochastic volatility model using deep learning methods for a fixed grid of strikes and maturities. We describe their framework by specialising to the case of European call options for a fixed stochastic volatility model.

The grid-based method consists of two stages: (i) learning and (ii) calibration. In the learning stage, one chooses a stochastic volatility model parametrised by a vector $\theta \in \Theta$,

where Θ is the admissible parameter space for the chosen stochastic volatility model. In addition, one fixes a finite set of strikes $K_1, \ldots, K_m$ and maturities $T_1, \ldots, T_n$ on which pricing will be performed. For each admissible vector θ, one then computes the Black-Scholes implied volatility surface corresponding to the choice of θ, which takes the form of a matrix $\sigma(\theta) \in \mathbb{R}^{m \times n}$.

Having produced a training dataset $\{(\theta_r, \sigma(\theta_r))\}_{r=1}^{N_{Train}} \subseteq \Theta \times \mathbb{R}^{m \times n}$, we proceed to learn the mapping $\theta \mapsto \sigma(\theta)$ using a neural network. In particular, we let

$$\begin{aligned} F_W : \Theta &\to \mathbb{R}^{m \times n} \\ \theta &\mapsto F_W(\theta) \end{aligned}$$

be a feedforward neural network with weights W. The weights W are chosen by solving the least-squares optimisation problem

$$\hat{W} = \arg\min_{W} \sum_{u=1}^{N_{\text{Train}}} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(F_W(\theta_u)_{ij} - \sigma_{BS}^{\mathcal{M}(\theta_u)}(K_i, T_j) \right)^2 \tag{3.13}$$

That is, we train a neural network to learn the mapping between a parameter set and a fixed set of points on the volatility surface.

The second step, calibration, assumes we have already fit a neural network $F_{\hat{W}}$. Suppose we have observed market implied volatilities on our discrete grid, $\sigma^{MKT} \in \mathbb{R}^{m \times n}$; then, calibrating the volatility model at hand is tantamount to solving the following least-squares optimisation problem:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left\| F_{\hat{W}}(\theta) - \sigma^{MKT} \right\|_2^2 = \arg\min_{\theta \in \Theta} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(F_{\hat{W}}(\theta)_{ij} - \sigma^{MKT}(K_i, T_j) \right)^2 \tag{3.14}$$

This optimisation problem can be solved by any standard least-squares solver, such as the Levenberg-Marquardt algorithm [60]. Note that option prices on the grid points can be extracted by means of the Black-Scholes equation, but not outside the grid points.

One of the key advantages of this method is that the analyst is in full control of the error in the approximating network. Indeed, if one produces the training dataset using Monte Carlo simulation, then a simple application of the strong law of large numbers tells us we can control the error in the simulated volatility surface to any extent that we want. Then, an application of the universal approximation theorem implies that the neural network can learn the mapping $\theta \mapsto \sigma(\theta)$ arbitrarily closely. A further advantage is that the slow

training step can be performed offline; once one has access to a fitted neural network, calibration takes a handful of milliseconds.

The main disadvantage, however, is that one needs to choose, *a priori*, a grid on which to sample the volatility surface. For real world applications, this can rapidly lead to model misspecification, as time decay can displace the time axis in the grid, and changing market liquidity as expiry dates approach may limit the availability of option data far in or out of the money. The authors of [45] argue that this can be resolved by performing arbitrage-free interpolation for data outside the pre-set grid. Indeed, this is a way to solve the issue, but in practice it would require constantly re-training models as expiries and strikes displace themselves from the original model grid. Moreover, this approach would prove counterproductive if one were trying to obtain a time series of the market's risk-neutral volatility parameters. Obtaining such time series, in particular for the Hurst parameter H, is important, as it allows us to line up abrupt changes in market parameters with economic events and thus provide a qualitative justification for our choice of stochastic volatility models.

One way to rectify this is to revert to a previous prevailing neural formulation, which instead learnt the mapping

$$\begin{aligned} \Theta \times \mathbb{R} \times \mathbb{R} &\to \mathbb{R} \\ (\theta, K, T) &\mapsto \sigma(\theta, K, T) \end{aligned}$$

Indeed, the universal approximation theorem guarantees us that we can fit a neural network that can learn such a mapping; it, however, does not guarantee an arbitrage-free mapping can be learnt. One could, in theory, modify their objective function to include arbitrage-free constraints, such as $\frac{\partial^2 C}{\partial K^2} > 0$. However, this introduces the extra problem of computing (higher-order) partial derivatives on a sparse grid, which could introduce numerical inaccuracy and could make training take longer.

A second way to account for out-of-grid volatilities is to augment the learning step by not just learning the pricing map, but also learning the interpolation map. In the next section, we expound on the theoretical framework to learn an interpolating scheme for volatility surfaces.

3.4.2 Functional neural volatility: An abstract framework for learning a volatility interpolation scheme

While in practice one only observes the volatility at a discrete set of points, and may wish to consider the surface as a finite-dimensional object, in theory the volatility surface is a

function. Thus, the right setting to study a volatility surface is a suitable function space. Hereafter, we assume that the volatility surface is function $\sigma : K \subseteq \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$. We treat $\sigma \in X$, where X is a suitable Banach space of functions.

Now, there is no need to restrict our study to the implied volatility surface. Any other sufficiently rich object that contains enough information about the stock price process to perform calibration is sufficient. For instance, one may instead choose the call or put option surface, or a risk-free density kernel for the stock price. Again, such objects are infinite-dimensional, so we study them as objects in a suitable Banach space. For simplicity, we call any such object a pricing map $\mathcal{P}$, or $\mathcal{P}_\theta$ when the parameter vector θ of a stochastic volatility model is specified.

Informally, we say that $\hat{\mathcal{P}}$ is an interpolation scheme for $\mathcal{P}$ if $\hat{\mathcal{P}} \in X$ as well, and $\hat{\mathcal{P}}$ is an arbitrage-free surface which is close in norm to $\mathcal{P}$. Several approximation schemes, such as cubic splines, or parametric representations of the volatility surface and the option price surface exist and are used in practice. Non-parametric schemes, such as radial-basis networks, exist and are used in practice as well—however, such methods, often encountered in data science applications, do not always guarantee an arbitrage-free fit.

Akin to the method introduced in Section 3.4.1, our approach also consists of two-steps: learning and calibration. The key difference is that in the learning step, instead of learning a grid representation for $\mathcal{P} \in \mathbb{R}^{m \times n}$, we instead learn a functional representation $\sigma \in X$. In particular, our learning phase proceeds as follows:

1. Fix a stochastic volatility model parametrised by $\theta \in \Theta$;
2. Produce a sufficiently rich grid on which the pricing map $\mathcal{P}$ is computed for a wide range of parameters θ;
3. Fit an interpolation scheme $\hat{\mathcal{P}}$ to $\mathcal{P}$;
4. Find a neural network representation F_W of the interpolation scheme $\hat{\mathcal{P}}$.

Once one concludes this learning step, the calibration step proceeds as in Equation (3.14), using a least-squares solver.

We call this method the "functional neural volatility" method, as it involves learning a representation in function space. This follows the tradition from the field of functional data analysis, where data oftentimes appears in the space of continuous functions or a suitable L^p space. For a detailed account of functional data analysis, we refer the reader to the textbook by Ramsay and Silverman [65].

This approach has several advantages over the approach proposed in [45]. First, one can push the interpolation step to the learning phase, where a slow algorithm does not interfere with fast model calibration. Secondly, since the neural network F_W has learnt an interpolation scheme, the neural network itself is now able to produce Black-Scholes implied volatilities outside of the grid on which the training set was produced. Thirdly, this network does not need to be re-trained when the market experiences time decay and the grid points shift; this allows us to use a single network to compute a time series of the stochastic volatility parameters the market experiences. Finally, if the interpolation scheme was truly arbitrage-free, then the neural network inherits that property automatically, which means we can perform arbitrage-free calibration across a wider range of expirations and strikes.

Now, we specify how one may learn $\mathcal{P}$ for different kinds of pricing maps. First, suppose $\mathcal{P}$ is simply the call option surface, C, where $C(K,T)$ is the call price at strike K and maturity T. While in the markets we observe C on a finite grid, elementary arbitrage theory gives us a set of simple conditions to ensure that an interpolating scheme $\hat{C}$ be arbitrage free. Namely, $\hat{C}(\cdot,T)$ must be convex, $\hat{C}(K,\cdot)$ must be non-decreasing, $\lim_{K\to\infty}\hat{C}(K,T)=0$ for all $T\geq 0$, $(S_0-K)^+\leq\hat{C}(T,K)\leq S_0$, and $\hat{C}(K,0)=(S_0-K)^+$. We may also require that C and $\hat{C}$ agree on the observed points. Given a finite set of grid points on which the call prices are free of arbitrage, there are several functions $\hat{C}$ that meet the conditions above. We may pick one representation (parametric or otherwise) for $\hat{C}$ and then proceed with step (4) of the learning phase outlined above. In particular, note that since an arbitrage-free representation for C is learnt, the functional neural volatility model is guaranteed to be arbitrage-free.

In the paragraph above, the call option price surface could easily be changed to the put option surface $P(K,T)$, with similar no-arbitrage conditions holding. Likewise, the surface could be replaced by the local or implied volatility surface, although in the case of implied volatility, no-arbitrage conditions are harder to write down.

Finally, given a grid of option prices, one may choose to learn a representation for the risk-neutral measure. For instance, in [12], the author represents the risk-neutral measure μ in terms of the discrete strike-derivatives of the call option surface. One may then take $\mathcal{P}=\mu$ and learn a representation of such a measure with the recipe above. This learnt representation can then be used to price options at any strike or maturity.

For concreteness, we specialise our abstract approach to a popular interpolation scheme for the implied volatility surface. Our choice to learn the implied volatility case is rooted on the fact that traders often think about the relative expensiveness of options in terms of implied volatility, which makes our proof-of-concept case more amenable to practitioners.

The surface stochastic volatility inspired interpolation scheme

Parametric formulas for the implied volatility surface are of the highest interest for practitioners, as they allow them to describe and explain the market with few parameters. A successful parametrisation of the volatility surface for fixed maturities is the stochastic volatility inspired (SVI) formula, introduced at Merrill Lynch in the late 90s by Gatheral, and later made available publicly in 2004 [34]. The SVI formula is parsimonious, as it consists only of five parameters; despite this, no simple no-arbitrage conditions for it are known.

Several researchers attempted to extend the SVI formula to fit the entire volatility surface. In 2012, Gatheral and Jacquier proposed an extension of the SVI formula, dubbed the surface SVI (or SSVI for short), which can fit the entire surface [36]. The SSVI formula is parametrised by the at-the-money total variance and an additional curvature function (which is usually parametrised by only three extra parameters). The SSVI model also enjoys the fact that there exist simple arbitrage-free conditions on its parameters.

Specifically, SSVI provides a parametrisation for the implied total variance curve. If one lets $k = \log \frac{K}{S_t}$ be the log-moneyness, T the maturity, and $\sigma(k,T)$ the Black-Scholes implied volatility at (k,T), then the total implied variance is given by

$$w(k,T) = \sigma^2(k,T)T \tag{3.15}$$

If we introduce the at-the-money implied total variance $\theta_T = \sigma^2(0,T)T$, then we may now define the SSVI formula.

Definition 3.3. Let $\varphi : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ be a C^∞ map such that $\lim_{T\to 0} \theta_T \varphi(\theta_T)$ exists in $\mathbb{R}$. The SSVI surface is then:

$$w(k,\theta_T) = \frac{\theta_T}{2}\left(1 + \rho\varphi(\theta_T)k + \sqrt{(\varphi(\theta_T)k+\rho)^2 + (1-\rho^2)}\right) \tag{3.16}$$

where $\rho \in [-1,1]$.

A popular choice for φ derives from the power law family; i.e. $\varphi(\theta_T) = \eta\theta_T^{-\lambda}$.

Now, we may specialise our functional neural volatility method to this choice of interpolation scheme. Below is a step-by-step learning procedure to fit this method to the SSVI parametrisation given above:

1. Fix a stochastic volatility model parametrised by $\theta \in \Theta$;

2. Produce a sufficiently rich grid on which the Black-Scholes implied volatility $\sigma(\theta)$ is computed;

3. Compute the at-the-money total variance curve and the implied total variance curves corresponding to $\sigma(\theta)$;

4. Given the model implied total variance curve produced in the preceding step, we run a least-squares solver to fit this curve using Equation (3.16) with $\varphi(\theta_T) = \eta\theta_T^{-\lambda}$, and extract the formula parameters ρ, η, λ;

5. Find a neural network representation F_W to fit the mapping $\theta \mapsto (\rho, \eta, \lambda)$.

The calibration step proceeds similarly to the grid-based method, with an additional step:

1. Given market observed implied volatilities, fit the SSVI parametrisation to them and recover the observed SSVI parameters $(\hat{\rho}, \hat{\eta}, \hat{\lambda})$;

2. Recover θ by solving the following least-squares optimisation problem:

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min} \left\| F_W(\theta) - (\hat{\rho}, \hat{\eta}, \hat{\lambda}) \right\|_2^2 \tag{3.17}$$

If the dimension of θ is bigger than 3 (as in the Heston model), then we may run into a problem: the inverse map would be underdetermined, so the recovered values of $\hat{\theta}$ may have a large error. This is not a major issue, as one can add the liquid at-the-money curve to the learning phase (i.e. we learn $(\rho, \eta, \lambda, \sigma(0, t))$, which, in practice, makes the problem well-posed.

3.4.3 Results on synthetic data

To test out our methods, we produced synthetic datasets using the rough Bergomi model. For each set of stochastic volatility parameters, we produced SSVI parameters. We fit a neural network mapping the stochastic volatility parameters to the SSVI parameters and at-the-money implied volatilities together. We then performed calibration via the Levenberg-Marquadt algorithm, as described in the previous section. In the results reported below, we report the relative errors for each stochastic volatility parameter, and report the mean and median relative errors for each parameter as well.

For the rough Bergomi model, we also assume that $\xi_0(t) = \xi_0$ in Equation (3.5). In addition, we used a simulated dataset for rough Bergomi implied volatilities whose parameter sets had the following ranges:

1. $0.01 \leq \xi_0 \leq 0.16$
2. $0.3 \leq \nu \leq 4$
3. $-0.95 \leq \rho \leq -0.1$
4. $0.02 \leq H \leq 0.5$

These parameter ranges were chosen as they are fairly conservative estimates of plausible market parameters that appear during calibration for each of the models.

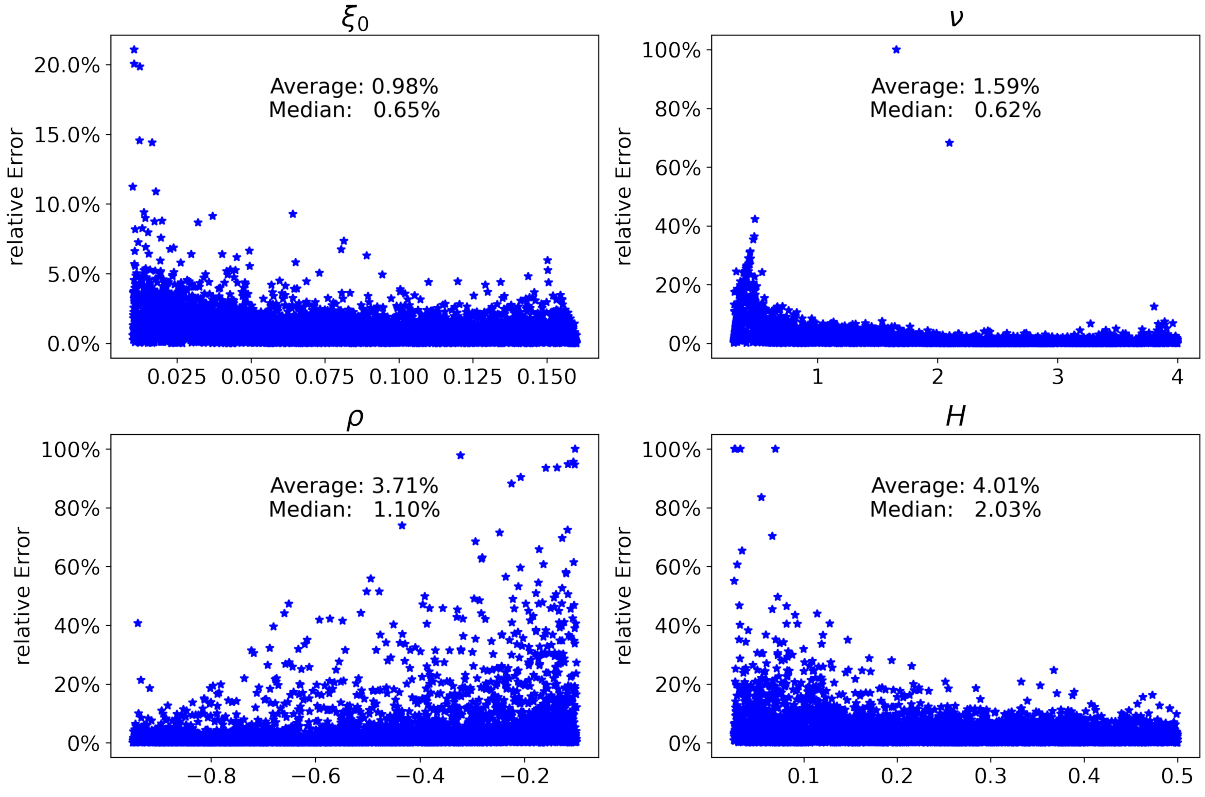

Figure 3.2: Relative errors in fitting procedure for rough Bergomi parameters

The errors in our parameter calibration are comparable to the errors observed in [45], with the advantage that our model can produce implied volatilities across all strikes and maturities in an arbitrage-free manner.

3.4.4 Universal approximation for functional neural networks

To finish our discussion on functional neural volatility networks, we wish to provide a theoretical justification for their use. Just like feed-forward neural networks enjoy universal approximation properties [20, 44], we wish to show that infinite-dimensional approximators also enjoy this property. The universal approximation theorem, first introduced by Cybenko, makes use of a certain measure-theoretic property that admissible activation functions enjoy, called the discriminatory property. Since Cybenko's result, several new universal approximation results have appeared in the deep learning literature. We refer the reader to [4, 15, 48, 50, 55, 64] and the references therein.

Since the class of functional neural networks we previously introduced act on function spaces, the right formalism to study universal approximation in function spaces needs to be introduced. We do so in this section by providing an extension to the methods in [20].

Definition 3.4. Fix a function $s : \mathbb{R} \to \mathbb{R}$ and a Banach space of functions X. A functional neural network in X with activation function s is a function:

$$N(x) = \sum_{j=1}^{N} \alpha_j s\left(f_{y_j}(x) + \theta_j\right) \qquad \alpha_j, \theta_j \in \mathbb{R} \text{ and } f_{y_j} \in X^* \tag{3.18}$$

where X^* is the dual space of X.

In the above definition, a neural network will have both N and its associated functionals $f_{y_1}, \ldots, f_{y_N}$ fixed. If we let them vary, then we consider instead the entire class of neural networks activated by s:

$$\mathcal{N}_s = \left\{ \sum_{j=1}^{N} \alpha_j \sigma\left(f_{y_j}(x) + \theta_j\right) \ : \ f_{y_j} \in X^*, \theta_j \in \mathbb{R}, N \in \mathbb{Z}^+ \right\} \tag{3.19}$$

A universal approximation theorem is a statement that shows that, under suitable restrictions, the set $\mathcal{N}_s$ is dense in a well-chosen function space. We need one more definition before proving such a theorem.

Definition 3.5. Let $s : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. We say that s is **functionally discriminatory** on a compact set $K \subseteq X$ if the only Radon measure $\mu \in M(K)$ such that the integral equation

$$\int_K s\left(f_y(x) + \theta\right) d\mu(x) = 0 \tag{3.20}$$

holds for all $f_y \in X^*$ and $\theta \in \mathbb{R}$ is the zero measure, $\mu = 0$.

Of course, proving a universal approximation theorem without regard for the structure of the function space X would be too ambitious. We may first prove this theorem for an interesting class of spaces, and then note a mild generalisation.

Theorem 3.1. Set $1 < p < \infty$. Let $s : \mathbb{R} \to \mathbb{R}$ be continuous and $K = B_{L^p} = \left\{ f \in L^p : \|f\|_p \leq 1 \right\}$ be the ball in L^p equipped with the weak-star topology. Define the set:

$$\mathcal{N}_s = \left\{ \sum_{j=1}^{N} \alpha_j s\left(f_{y_j}(x) + \theta_j\right) : f_{y_j} \in L^p[0,1]^*, \theta_j \in \mathbb{R}, N \in \mathbb{Z}^+ \right\} \tag{3.21}$$

Then, s is a functionally discriminatory function if and only if $\overline{\mathcal{N}_s} = C\left(K\right)$.

Proof. Letting q be the Hölder conjugate of p, all continuous linear functionals f_{y_j} are represented by

$$f_{y_j}(x) = \int_{[0,1]} x(t) y_j(t) d\lambda(t) \qquad y_j \in L^q[0,1] \tag{3.22}$$

By the Banach-Alaoglu theorem, the set K is weak-star compact; since L^q is separable, the weak-star topology on K is metrisable, so we may regard K as a compact metric space. Observe that the functionals $f_{y_j} : K \to \mathbb{R}$ are weak-star continuous, so we conclude that $\mathcal{N}_s \subseteq C(K)$.

($\Longleftarrow$) Arguing by contradiction, suppose $\overline{\mathcal{N}_s} \neq C\left(K\right)$. By the Hahn-Banach theorem, there exists a continuous functional $F : C(K) \to \mathbb{R}$ such that $F \neq 0$ but $F(x) = 0$ for all $x \in \overline{\mathcal{N}_s}$. Since K is compact, the Riesz Representation theorem tells us that for all $x \in C(K)$ we have:

$$F(x) = \int_K x(t) d\mu(t) \tag{3.23}$$

For a Radon measure $\mu \in M(K)$. In particular, $s(f_y(x)+\theta) \in \overline{\mathcal{N}_s}$ for all $f_y \in L^p[0,1]^*$ and $\theta \in \mathbb{R}$, thus

$$F(x) = \int_K s(f_y(x(t)) + \theta) d\mu(t) \tag{3.24}$$

for all f_y and θ. Since s is a functional discriminatory function, this implies that $\mu = 0$, contradicting the fact that $F \neq 0$. Hence $\overline{\mathcal{N}_s} = C(K)$

($\Longrightarrow$) $\mathcal{N}_s$ is a normed vector space. If $\mathcal{N}_s$ is dense in $C(K)$, then the continuous linear functionals on $\mathcal{N}_s$ must extend uniquely to all of $C(K)$. Since the zero functional extends uniquely to the zero functional on $C(K)$, the Riesz Representation Theorem implies that this functional corresponds to integration against the zero measure. Hence, s is discriminatory. □

In particular, the proof above achieves an extension of Theorem 1 in [20] by introducing the necessary formalism for an infinite-dimensional universal approximation theorem. Such extension is necessary to justify the use of neural networks for functional data. Note that the key in achieving this extension is that the set K above is a metrisable compact space. The only requirement to achieve this is that the ambient space X be separable and reflexive, as is the case with $L^p[0,1]$ for $1 < p < \infty$.

3.5 Deep learning H under the empirical measure

One of the key observations when modelling a volatilities using models driven by fractional Brownian motion is that the roughness of the paths does not change under equivalent change in measure. That is, for several classes of models, we can make a statement of the form "under the physical measure, volatility v_t is $H - \epsilon$ Hölder continuous"; since equivalent measures agree on which measurable events have full probability, this statement holds true if and only if we can say "under the risk-neutral measure, volatility v_t is $H - \epsilon$ Hölder continuous".

Thus, if we are interested in detecting the time series of roughness of a volatility trajectory, we could measure it either in risk-neutral or in physical space. The previous sections were concerned with modelling roughness in risk-neutral space; this section focuses on measuring roughness of the physical paths of volatility. In particular, we focus on discussing the roughness of models more general than Gatheral's rough fractional stochastic volatility model, which is driven by fractional Brownian motion.

3.5.1 A more general rough volatility model: bifractional Brownian motion

The bifractional Brownian motion, introduced by Houdré and Villa [46], is a continuous, centred Gaussian process $B_t^{H,K}$, whose covariance is given by:

$$R^{K,H}(s,t) = \frac{1}{2^K}\left[(t^{2H} + s^{2H})^K - |t-s|^{2HK}\right] \tag{3.25}$$

and $B_0^{H,K} = 0$, where $H \in (0,1)$ and $K \in (0,1]$. Note that when $K = 1$, this reduces to a fractional Brownian motion, so that a bifractional Brownian motion is indeed a generalisation of the fractional Brownian motion. Some interesting properties for bifractional Brownian motion are [67, 74]:

1. For small $|t-s|$, it is "almost stationary"; more precisely:
$$2^{-K}|t-s|^{2HK} \leq \mathbb{E}\left(B_t^{H,K} - B_s^{H,K}\right) \leq 2^{1-K}|t-s|^{2HK}$$
2. The process is HK-self-similar
3. The process is almost surely $HK - \epsilon$-Hölder continuous, for $\epsilon > 0$.
4. If $HK = \frac{1}{2}$, the process admits linear quadratic variation, but it is a semimartingale if and only if it is a Brownian motion.

Just like we regarded H as the roughness parameter for fractional Brownian motion, we may analogously regard the product HK as the roughness for bifractional Brownian motion. In what comes, we study the roughness of volatility under the assumption that the log-volatility process is driven by a bifractional Brownian motion. To do so, we rely heavily on the deep learning methods to study the roughness of time series first developed by Stone [72]. In particular, Stone makes use of convolutional neural networks to learn the Hurst exponent of the following Gaussian process, which is one of the building blocks of the rough Bergomi model:

$$\int_0^t (t-s)^{H-1/2} dW_s \tag{3.26}$$

We adapt these methods to learn the Hurst exponent of fractional Brownian motion directly and to learn the product HK for bifractional Brownian motion. We use the built neural networks to then construct time series for the roughness of the volatility processes of several financial assets. In particular we train two networks: one trained solely on fractional Brownian motion data and another trained on bifractional Brownian motion data.

3.5.2 Volatility roughness under bifractional Brownian motion

We gathered daily realised volatility proxies for the assets available in the Oxford-Man Institute volatility dataset[1]. Their library includes several measures of realised volatility and can be downloaded as a zip file with data going back to the year 2000 for several major assets. These data were loaded into Python for the analysis pertaining to this thesis. For each dataset, we use 100 contiguous days to estimate the roughness of the time series; that is to estimate the Hurst parameter at time t, we look at the time series

$$s_k^t = \log(\sigma_{t+k}) - \log(\sigma_t) \qquad k = 1, 2, \ldots, 100 \tag{3.27}$$

We then feed the time series $(s_1^t, s_2^t, \ldots, s_{100}^t)$ to our neural networks to get the value for H_t; we do the same to obtain the value of $(HK)_t$.

Our plots for the time series of H_t and $(HK)_t$ for three assets of interest are shown in the figure below; statistics for the time series of several more assets are shown in the tables following our figures.

[1]This methodology is used in the dataset made publicly available by the Oxford-Man Institute: https://realized.oxford-man.ox.ac.uk/

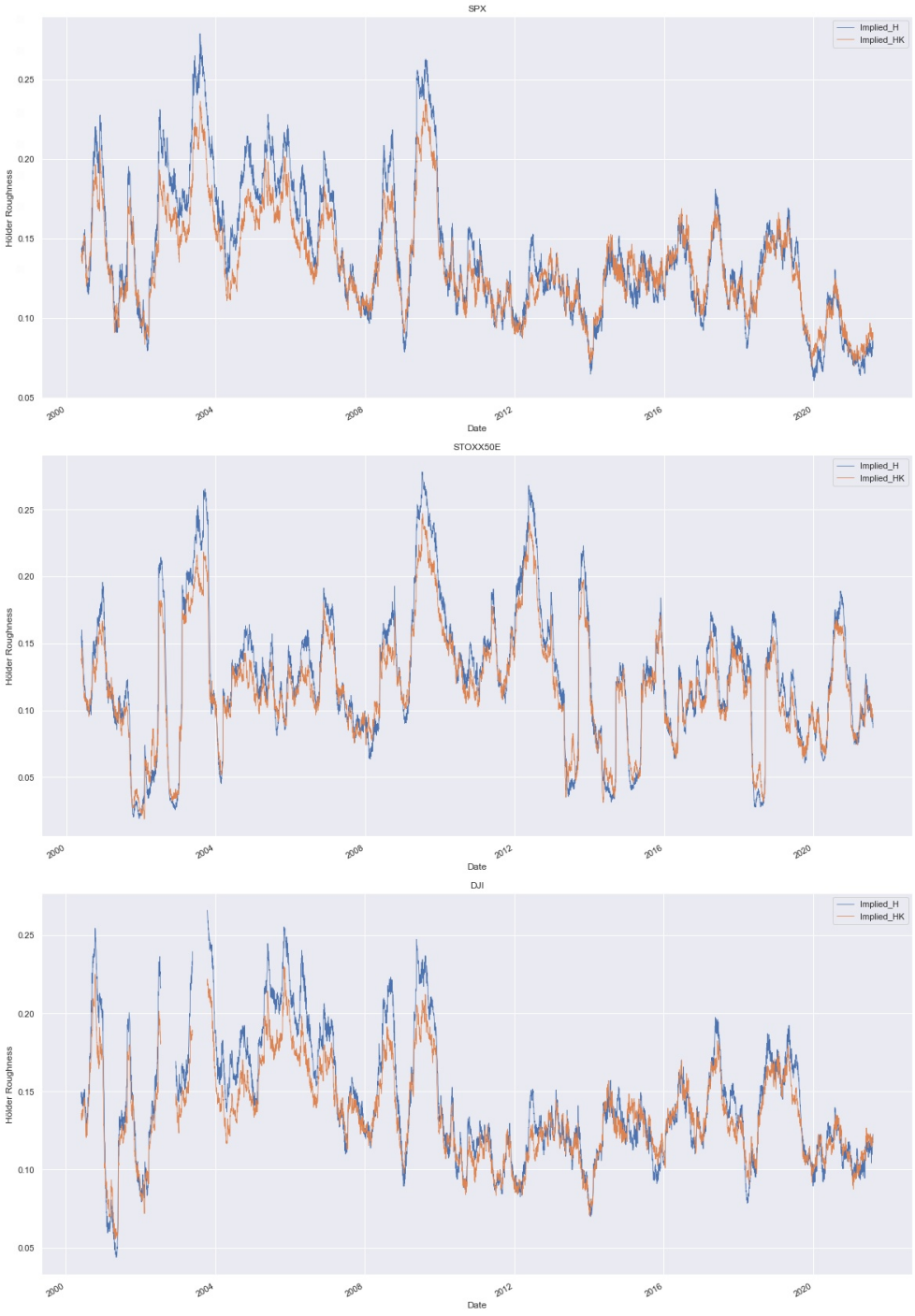

Figure 3.3: H_t and $(HK)_t$ for the volatility series for the SPX, EURO STOXX 50, and Dow Jones Industrial average indices.

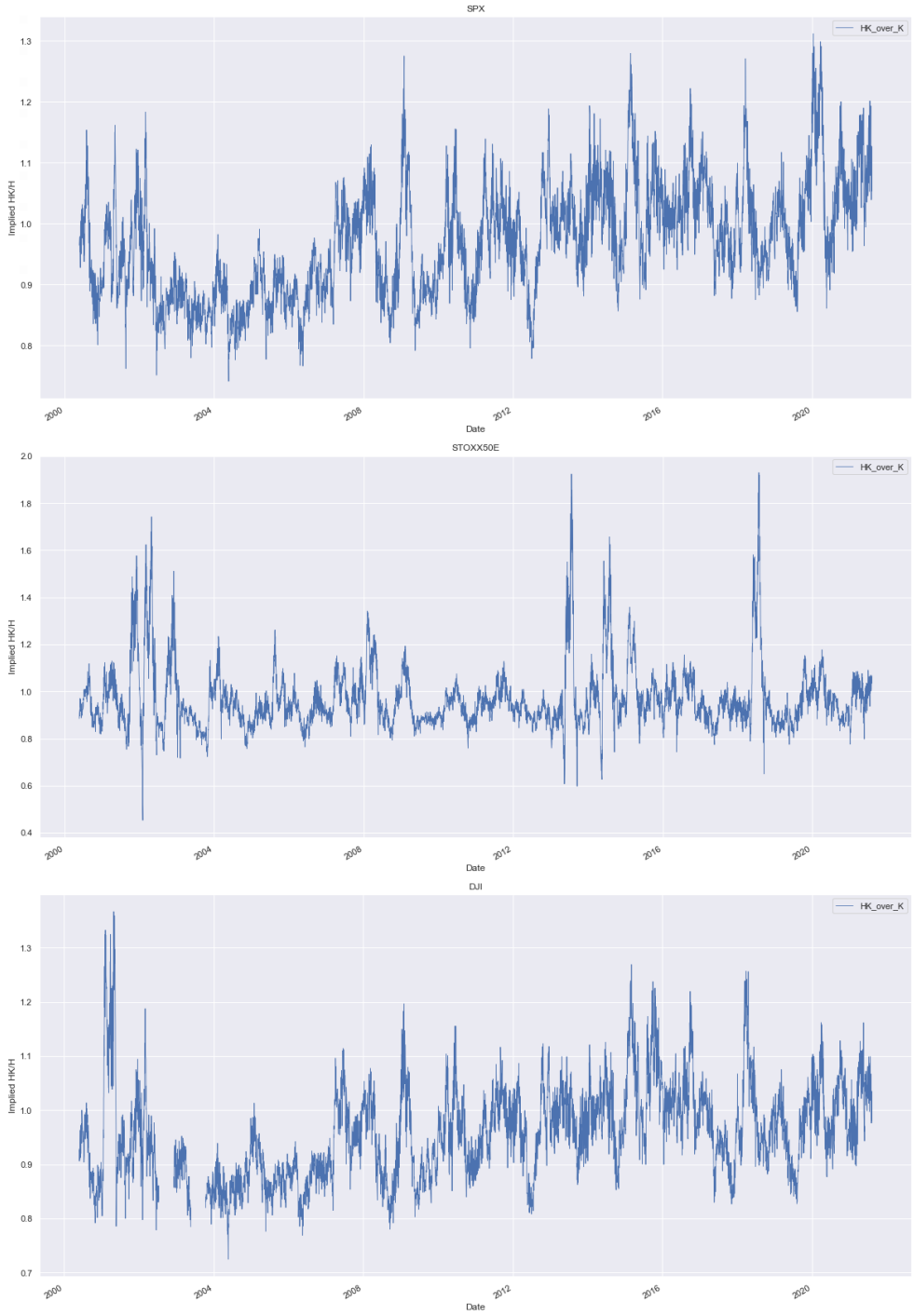

Figure 3.4: The ratio $\frac{(HK)_t}{H_t}$ for the volatility series for the SPX, EURO STOXX 50, and Dow Jones Industrial average indices.

Security	min	max	median	mean
AEX	0.072	0.287	0.155	0.159
AORD	0.047	0.214	0.141	0.135
BFX	0.076	0.314	0.178	0.182
BSESN	0.058	0.280	0.153	0.155
BVLG	0.109	0.287	0.208	0.202
BVSP	0.054	0.244	0.140	0.142
DJI	0.044	0.266	0.134	0.143
FCHI	0.061	0.284	0.152	0.155
FTMIB	0.082	0.275	0.178	0.174
FTSE	0.065	0.252	0.141	0.144
GDAXI	0.037	0.303	0.159	0.163
GSPTSE	0.051	0.241	0.128	0.130
HSI	0.091	0.320	0.202	0.200
IBEX	0.064	0.330	0.176	0.178
IXIC	0.067	0.298	0.156	0.156
KS11	0.095	0.285	0.175	0.177
KSE	0.005	0.213	0.102	0.100
MXX	0.038	0.269	0.150	0.151
N225	0.056	0.240	0.149	0.146
NSEI	0.024	0.260	0.137	0.136
OMXC20	0.027	0.291	0.157	0.156
OMXHPI	0.037	0.241	0.137	0.137
OMXSPI	0.041	0.239	0.139	0.141
OSEAX	0.069	0.269	0.165	0.166
RUT	0.005	0.225	0.109	0.109
SMSI	0.049	0.247	0.138	0.141
SPX	0.061	0.279	0.132	0.141
SSEC	0.019	0.311	0.154	0.149
SSMI	0.069	0.364	0.254	0.253
STI	0.024	0.371	0.253	0.250
STOXX50E	0.019	0.278	0.123	0.125

Table 3.1: Statistics for the implied H_t time series for each stock

Security	min	max	median	mean
AEX	0.071	0.242	0.141	0.145
AORD	0.052	0.187	0.127	0.123
BFX	0.081	0.265	0.159	0.163
BSESN	0.061	0.250	0.144	0.145
BVLG	0.103	0.256	0.184	0.179
BVSP	0.065	0.211	0.134	0.133
DJI	0.055	0.230	0.131	0.135
FCHI	0.066	0.252	0.140	0.141
FTMIB	0.083	0.237	0.162	0.159
FTSE	0.073	0.214	0.128	0.131
GDAXI	0.037	0.259	0.147	0.148
GSPTSE	0.056	0.217	0.121	0.123
HSI	0.085	0.261	0.174	0.174
IBEX	0.066	0.281	0.160	0.161
IXIC	0.076	0.269	0.147	0.147
KS11	0.096	0.250	0.158	0.159
KSE	0.013	0.186	0.100	0.098
MXX	0.044	0.235	0.135	0.137
N225	0.055	0.206	0.137	0.136
NSEI	0.028	0.231	0.130	0.128
OMXC20	0.035	0.254	0.141	0.141
OMXHPI	0.049	0.217	0.130	0.128
OMXSPI	0.049	0.209	0.129	0.130
OSEAX	0.072	0.239	0.149	0.150
RUT	0.011	0.207	0.106	0.106
SMSI	0.052	0.220	0.131	0.132
SPX	0.070	0.237	0.128	0.133
SSEC	0.032	0.267	0.143	0.139
SSMI	0.068	0.319	0.222	0.220
STI	0.040	0.367	0.218	0.217
STOXX50E	0.019	0.247	0.115	0.117

Table 3.2: Statistics for the implied $(HK)_t$ time series for each stock

Security	min	max	median	mean
AEX	0.737	1.153	0.910	0.918
AORD	0.704	1.272	0.908	0.916
BFX	0.731	1.151	0.895	0.903
BSESN	0.749	1.337	0.936	0.946
BVLG	0.736	1.032	0.884	0.886
BVSP	0.725	1.331	0.937	0.946
DJI	0.725	1.367	0.952	0.958
FCHI	0.739	1.193	0.914	0.923
FTMIB	0.743	1.195	0.913	0.919
FTSE	0.715	1.267	0.913	0.923
GDAXI	0.608	1.260	0.913	0.918
GSPTSE	0.689	1.230	0.948	0.956
HSI	0.721	1.105	0.869	0.876
IBEX	0.706	1.177	0.904	0.913
IXIC	0.744	1.324	0.948	0.959
KS11	0.744	1.176	0.899	0.906
KSE	0.696	3.415	0.987	1.018
MXX	0.748	1.463	0.907	0.927
N225	0.729	1.370	0.928	0.943
NSEI	0.717	1.344	0.946	0.960
OMXC20	0.781	1.542	0.909	0.939
OMXHPI	0.768	1.419	0.938	0.951
OMXSPI	0.742	1.362	0.920	0.935
OSEAX	0.737	1.255	0.903	0.915
RUT	0.535	2.658	0.981	0.996
SMSI	0.746	1.310	0.944	0.954
SPX	0.741	1.312	0.964	0.969
SSEC	0.758	2.064	0.938	0.976
SSMI	0.749	1.139	0.869	0.875
STI	0.676	7.315	0.856	0.909
STOXX50E	0.453	1.930	0.939	0.970

Table 3.3: Statistics for the ratio $\frac{(HK)_t}{H_t}$ time series for each stock

For all of the considered stocks that the time series for H_t and $(HK)_t$ are qualitatively

similar, giving credence to the belief that the log-volatility series can be modelled by a fractional Brownian motion with $H \approx 0.1$. This is a surprising finding, as one would expect that under greater model flexibility, the roughness of the trajectory would be different from the observed under the assumption that log-volatility is a fractional Brownian motion.

3.5.3 Transfer learning of processes

We began this chapter by arguing that Knightian risk is a reason to be cautious when specifying a market model. Yet, we proceeded by considering only a slightly more general class of models as a means to immunise us against model misspecification. One could argue that this does not confer enough protection against Knightian risk, as there are still several classes of models that we have not learnt using the methods in this chapter. However, not all is lost. In the remainder of this chapter we provide empirical and theoretical evidence for why learning the roughness of one process is sufficient to calculate the roughness of a different process.

Definition 3.6. Let N_t be a stochastic process whose paths are almost surely locally L^1; i.e. for almost all ω we have that $N_t(\omega)$ is Borel measurable in t and $\int_0^t |N_s(\omega)|ds < \infty$ for all $t \geq 0$. We say that X_t is an Ornstein-Uhlenbeck process driven by N_t if it is the solution (in some sense) to the stochastic differential equation

$$dX_t^\alpha = -\alpha(X_t - m)dt + \lambda dN_t \tag{3.28}$$

where $\alpha > 0$ is a parameter controlling the mean-reversion strength of X_t^α.

Following Appendix A from [14], we note that the solution to Equation (3.28) exists as a pathwise integral whenever the paths of N_t are almost surely continuous, as a simple consequence of the integration by parts formula. In such a case, we may interpret the unique pathwise solution of Equation (3.28) as

$$X_t^\alpha = m + \lambda \int_{-\infty}^t e^{-\alpha(t-s)} dN_s = m + \lambda N_t + \alpha\lambda \int_{-\infty}^t e^{-\alpha(t-s)} N_s ds \tag{3.29}$$

We notice that if α is small, then the exponential kernel is close to being flat on small intervals around zero; in such a regime, the increment $X_t^\alpha - X_0^\alpha$ is proportional to N_t. We formalise this below.

Theorem 3.2. Suppose that N_t in Equation (3.28) satisfies a maximal inequality of the form $\mathbb{E}\left(\sup_{t\in[0,T]}|N_t|\right) < \infty$.

Then,

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\alpha - X_0^\alpha - \lambda N_t|\right) \to 0$$

as $\alpha \to 0$.

Proof. Using the integration by parts formula in 3.29, we get

$$\begin{aligned} X_t^\alpha - X_0^\alpha - \lambda N_t &= -\alpha\lambda \int_{-\infty}^{t} e^{-\alpha(t-s)} N_s ds + \alpha\nu \int_{-\infty}^{0} e^{\alpha s} ds \\ &= -\alpha\lambda \int_{0}^{t} e^{-\alpha(t-s)} N_s ds + \alpha\lambda \int_{-\infty}^{0} \left(e^{\alpha s} - e^{-\alpha(t-s)}\right) N_s ds \end{aligned}$$

By the triangle inequality,

$$\sup_{t\in[0,T]}|X_t^\alpha - X_0^\alpha - \lambda N_t| \le \alpha\lambda T\tilde{N}_t + \alpha\nu \int_{-\infty}^{0} \left(e^{\alpha s} - e^{-\alpha(T-s)}\right) \tilde{N}_s ds$$

Where $\tilde{N}_t = \sup_{s\in[0,t]}|N_s|$. Taking expectations, using Tonelli's theorem, and applying the maximal inequality for N_t, we get:

$$\begin{aligned} \mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\alpha - X_0^\alpha - \lambda N_t|\right) &\le \alpha\lambda T\mathbb{E}\left(\tilde{N}_T\right) + \alpha\nu\mathbb{E}\left(\tilde{N}_T\right) \underbrace{\int_{-\infty}^{0} \left(e^{\alpha s} - e^{-\alpha(T-s)}\right) ds}_{<\infty} \\ &\to 0 \end{aligned}$$

as $\alpha \to 0$. □

Corollary 3.1. Let N_t be a centred Gaussian process with almost sure continuous paths. If X_t^α is an Ornstein-Uhlenbeck process driven by N_t, then

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\alpha - X_0^\alpha - \lambda N_t|\right) \to 0$$

as $\alpha \to 0$.

Proof. Since N_t has continuous paths, X_t^α can be defined pathwise. The conclusion of the previous theorem holds if N_t satisfies a maximal inequality in $[0, T]$, which happens as a consequence of Borell's inequality (cf. Theorem 2.1.1 in [1]). □

Naturally, bifractional (and hence fractional) Brownian motion satisfies the conditions in the previous corollary, so the conclusion of the theorem holds for bifractional Ornstein-Uhlenbeck processes. If someone, or a neural network, were observing the paths of bifractional Ornstein-Uhlenbeck processes with small α, then those paths would be difficult to distinguish from the paths of bifractional Brownian motion. By a similar token, such observer would be unable to distinguish the roughness of the processes. We evince this data feature below.

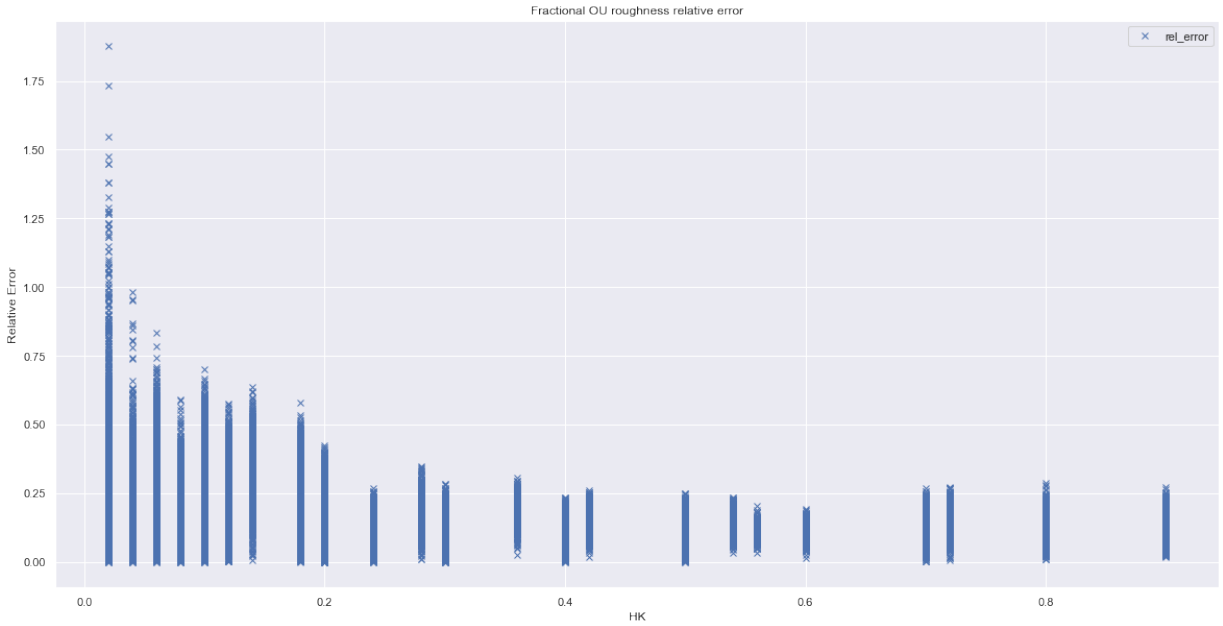

Figure 3.5: Relative errors for the model implied roughness plotted against true Hölder roughness HK

For most of the values for HK, the mean relative error is around 10%, except for the small HK case, where relative errors are magnified. That said, Figure 3.5 shows that the model's predictions for HK coincide with the true magnitude of HK; this serves as empirical evidence for the conclusion of Theorem 3.2.

References

[1] Robert J Adler and Jonathan E Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.

[2] Pieter C Allaart and Kiko Kawamura. The takagi function: a survey. *Real Analysis Exchange*, 37(1):1–54, 2012.

[3] Elisa Alòs and Kenichiro Shiraya. Estimating the hurst parameter from short term volatility swaps: a malliavin calculus approach. *Finance and Stochastics*, 23(2):423–447, 2019.

[4] Maximilian Baader, Matthew Mirman, and Martin Vechev. Universal approximation with certified networks. *arXiv preprint arXiv:1909.13846*, 2019.

[5] Andrew R Barron. Approximation and estimation bounds for artificial neural networks. *Machine learning*, 14(1):115–133, 1994.

[6] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[7] Lorenzo Bergomi. *Stochastic volatility modeling*. CRC press, 2015.

[8] Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic calculus for fractional Brownian motion and applications*. Springer Science & Business Media, 2008.

[9] Klaus Bichteler et al. Stochastic integration and l^p-theory of semimartingales. *The Annals of Probability*, 9(1):49–89, 1981.

[10] Avi Bick and Walter Willinger. Dynamic spanning without probabilities. *Stochastic processes and their Applications*, 50(2):349–374, 1994.

[11] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, 1973.

[12] Hans Buehler. Expensive martingales. *Quantitative Finance*, 6(3):207–218, 2006.

[13] Peter Carr and Dilip Madan. Option valuation using the fast fourier transform. *Journal of computational finance*, 2(4):61–73, 1999.

[14] Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. Fractional ornstein-uhlenbeck processes. *Electronic Journal of probability*, 8:1–14, 2003.

[15] Kai Fong Ernest Chong. A closer look at the approximation capabilities of neural networks. *arXiv preprint arXiv:2002.06505*, 2020.

[16] Rama Cont and Purba Das. Quadratic variation and quadratic roughness. *arXiv preprint arXiv:1907.03115*, 2019.

[17] Rama Cont and David-Antoine Fournié. Change of variable formulas for non-anticipative functionals on path space. *Journal of Functional Analysis*, 259(4):1043–1072, 2010.

[18] Rama Cont and Nicolas Perkowski. Pathwise integration and change of variable formulas for continuous paths with arbitrary regularity. *arXiv preprint arXiv:1803.09269*, 2018.

[19] Rama Cont and Candia Riga. Robustness and pathwise analysis of hedging strategies for path dependent derivatives. Technical report, Working paper, 2014.

[20] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[21] Mark Davis, Jan Obłój, and Pietro Siorpaes. Pathwise stochastic calculus with local times. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 54, pages 1–21. Institut Henri Poincaré, 2018.

[22] Claude Dellacherie. Un survol de la théorie de l'intégrale stochastique. In *Measure Theory Oberwolfach 1979*, pages 365–395. Springer, 1980.

[23] Bruno Dupire. Functional itô calculus. *Quantitative Finance*, 19(5):721–729, 2019.

[24] Bruno Dupire et al. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[25] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[26] Ronen Eldan and Ohad Shamir. The power of depth for feedforward neural networks. In *Conference on learning theory*, pages 907–940. PMLR, 2016.

[27] Hans Föllmer. Calcul d'itô sans probabilités. In *Séminaire de Probabilités XV 1979/80*, pages 143–150. Springer, 1981.

[28] Hans Föllmer, Philip Protter, and Albert N Shiryayev. Quadratic covariation and an extension of itô's formula. *Bernoulli*, pages 149–169, 1995.

[29] Hans Föllmer and Alexander Schied. Probabilistic aspects of finance. *Bernoulli*, 19(4):1306–1326, 2013.

[30] Martin Forde, Masaaki Fukasawa, Stefan Gerhold, and Benjamin Smith. The rough bergomi model as $h \to 0$–skew flattening/blow up and non-gaussian rough volatility. *preprint*, 2020.

[31] Peter K Friz and Martin Hairer. *A course on rough paths*. Springer, 2020.

[32] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. Is volatility rough? *arXiv preprint arXiv:1905.04852*, 2019.

[33] Matthieu Garcin. A comparison of maximum likelihood and absolute moments for the estimation of hurst exponents in a stationary framework. *arXiv preprint arXiv:2003.02566*, 2020.

[34] Jim Gatheral. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. *Presentation at Global Derivatives & Risk Management, Madrid*, page 0, 2004.

[35] Jim Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, 2011.

[36] Jim Gatheral and Antoine Jacquier. Arbitrage-free svi volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

[37] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[38] EG Gladyshev. A new limit theorem for stochastic processes with gaussian increments. *Theory of Probability & Its Applications*, 6(1):52–61, 1961.

[39] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[40] Patrick S Hagan, Deep Kumar, Andrew S Lesniewski, and Diana E Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.

[41] Xiyue Han. A short discussion on volatility is rough. unpublished, 2020.

[42] Sebas Hendriks and Claude Martini. The extended ssvi volatility surface. *Available at SSRN 2971502*, 2017.

[43] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[44] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[45] Blanka Horvath, Aitor Muguruza, and Mehdi Tomas. Deep learning volatility. *Available at SSRN 3322085*, 2019.

[46] Christian Houdré and José Villa. An example of infinite dimensional quasi-helix. *Contemporary Mathematics*, 336:195–202, 2003.

[47] John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The journal of finance*, 42(2):281–300, 1987.

[48] Ziwei Ji, Matus Telgarsky, and Ruicheng Xian. Neural tangent kernels, transportation mappings, and universal approximation. *arXiv preprint arXiv:1910.06956*, 2019.

[49] Ioannis Karatzas and Donghan Kim. Trading strategies generated pathwise by functions of market weights. *Finance and Stochastics*, 24(2):423–463, 2020.

[50] Patrick Kidger and Terry Lyons. Universal approximation with deep narrow networks. In *Conference on Learning Theory*, pages 2306–2327. PMLR, 2020.

[51] Donghan Kim. Local times for continuous paths of arbitrary regularity. *arXiv preprint arXiv:1904.07327*, 2019.

[52] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[53] Joachim Lebovits and Jacques Lévy Vehel. White noise-based stochastic calculus with respect to multifractional brownian motion. *Stochastics An International Journal of Probability and Stochastic Processes*, 86(1):87–124, 2014.

[54] Giulia Livieri, Saad Mouti, Andrea Pallavicini, and Mathieu Rosenbaum. Rough volatility: evidence from option prices. *IISE transactions*, 50(9):767–776, 2018.

[55] Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. The expressive power of neural networks: A view from the width. In *Advances in neural information processing systems*, pages 6231–6239, 2017.

[56] Benoit B Mandelbrot, Adlai J Fisher, and Laurent E Calvet. A multifractal model of asset returns. 1997.

[57] Benoit B Mandelbrot and John W Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[58] Yuliya Mishura and Alexander Schied. Constructing functions with prescribed pathwise quadratic variation. *Journal of Mathematical Analysis and Applications*, 442(1):117–137, 2016.

[59] Yuliya Mishura and Alexander Schied. On (signed) takagi–landsberg functions: pth variation, maximum, and modulus of continuity. *Journal of Mathematical Analysis and Applications*, 473(1):258–272, 2019.

[60] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[61] Eyal Neuman and Mathieu Rosenbaum. Fractional brownian motion with zero hurst parameter: a rough volatility viewpoint. *Electronic Communications in Probability*, 23:1–12, 2018.

[62] Ivan Nourdin. *Selected aspects of fractional Brownian motion*, volume 4. Springer, 2012.

[63] David Nualart. *The Malliavin calculus and related topics*, volume 1995. Springer, 2006.

[64] Allan Pinkus. Approximation theory of the mlp model in neural networks. *Acta numerica*, 8(1):143–195, 1999.

[65] James Ramsay and Bernard W Silverman. *Functional data analysis (Springer series in statistics)*. 1997.

[66] Johannes Ruf and Weiguan Wang. Neural networks for option pricing and hedging: a literature review. *Journal of Computational Finance, Forthcoming*, 2020.

[67] Francesco Russo and Ciprian A Tudor. On bifractional brownian motion. *Stochastic Processes and their applications*, 116(5):830–856, 2006.

[68] Alexander Schied. Model-free cppi. *Journal of Economic Dynamics and Control*, 40:84–94, 2014.

[69] Alexander Schied. On a class of generalized takagi functions with linear pathwise quadratic variation. *Journal of Mathematical Analysis and Applications*, 433(2):974–990, 2016.

[70] Alexander Schied, Leo Speiser, and Iryna Voloshchenko. Model-free portfolio theory and its functional master formula. *SIAM Journal on Financial Mathematics*, 9(3):1074–1101, 2018.

[71] Alexander Schied and Mitja Stadje. Robustness of delta hedging for path-dependent options in local volatility models. *Journal of applied probability*, 44(4):865–879, 2007.

[72] Henry Stone. Calibrating rough volatility models: a convolutional neural network approach. *Quantitative Finance*, 20(3):379–392, 2020.

[73] Tetsuya Takaishi. Rough volatility of bitcoin. *Finance Research Letters*, 32:101379, 2020.

[74] Ciprian A Tudor and Yimin Xiao. Sample path properties of bifractional brownian motion. *Bernoulli*, 13(4):1023–1052, 2007.

[75] M Wuermli. *Lokalzeiten für martingale*. PhD thesis, Master's thesis, Universität Bonn, 1980, supervised by Hans Föllmer, 1980.

[76] Laurence C Young. An inequality of the hölder type, connected with stieltjes integration. *Acta Mathematica*, 67(1):251, 1936.