IMES DISCUSSION PAPER SERIES

A Survey of Rough Volatility

Kazuhiro Hiraki and Yuji Shinozaki

Discussion Paper No. 2024-E-6

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

2-1-1 NIHONBASHI-HONGOKUCHO
CHUO-KU, TOKYO 103-8660
JAPAN

IMES Discussion Paper Series 2024-E-6
July 2024

A Survey of Rough Volatility

Kazuhiro Hiraki* and Yuji Shinozaki**

Abstract

Volatility, which is a measure of asset price uncertainty, is an important subject of finance research. In recent years, a growing number of studies have suggested that volatility is rough, namely that volatility fluctuates more wildly than is conventionally recognized, making rough volatility a new trend in volatility research. This strand of literature includes empirical analysis of the volatility observable in high-frequency trading data and derivatives markets, theoretical studies that clarify the necessity of rough volatility models, and their applications to derivative pricing. There have also been studies that explore the mechanism giving rise to roughness. In this paper, we provide an overview of the literature and the history of developments in the study of volatility from three perspectives—(1) time-series analysis of asset price fluctuations, (2) derivative pricing and risk management, and (3) the economic mechanism that gives rise to rough volatility—thereby clarifying the current position and prospects of rough volatility research. In particular, referring to issues in financial practice and relationship between volatility research and economics, we emphasize the importance of cross-disciplinary studies with volatility as a nexus.

Keywords: Historical volatility; Implied volatility; Market microstructure; Highfrequency trading; Derivative pricing; Risk management; Market forecast
JEL classification: D53, G12, G13, G17

* Director, Institute for Monetary and Economic Studies, Bank of Japan (currently, International Monetary Fund, E-mail:khiraki@imf.org)
** Deputy Director, Institute for Monetary and Economic Studies, Bank of Japan (currently, Associate Professor, Musashino University, E-mail:y-shino@musashino-u.ac.jp)

The authors appreciate comments from and discussions with Masaaki Fukasawa (Osaka University), Kazuhiko Ohashi (Hitotsubashi University and Tokyo Institute of Technology), and Toshiaki Watanabe (Hitotsubashi University). The authors also thank FORTE Science Communications (https://www.forte-science.co.jp/) for English language editing. The views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan, the IMF, its Executive Board, or IMF management.

Contents

1 Introduction

One of the key research themes in finance research is how uncertainty affects various financial issues, such as price formation mechanisms of financial assets, optimal investment behaviors, and risk management methods. Consequently, volatility, which is a measure of asset price uncertainty (i.e., magnitude of fluctuations), has been one of the most fundamental research subjects in finance, and research on volatility has developed in tandem with various adjacent fields, including economics, mathematical science, and computer science.

Volatility can be categorized into two types, depending on how it is measured: *historical volatility* (HV), which is a backward-looking measure of volatility calculated from past price fluctuations, and *implied volatility* (IV), which is a forward-looking measure of underlying assets' volatility calculated from derivative prices. Research on HV focuses on time-series data and has been developing volatility forecasting models and applications to investment strategies, whereas research on IV has been providing the theoretical foundation for pricing and hedging derivatives as well as for volatility indices such as the VIX index.

While the HV research and IV research differ in terms of the motivation for the analysis and how volatility is measured, they share the common goals of revealing stylized features of volatility dynamics (e.g., stochastic fluctuations) and developing realistic volatility models that express these features. In particular, both strands of research have been utilizing Brownian motions and random walks (the discrete-time counterparts of Brownian motions) as basic building blocks for modeling the stochastic dynamics of volatility.

From the 2000s onward, however, researchers have been reporting the *roughness* of volatility observed from IV and HV, which is considered difficult to reconcile with conventional Brownian motion-based models. Here, roughness refers to the time-series feature that a stochastic process fluctuates wildly in a short period of time, and a fractional Brownian motion is known to have this feature.[1] In the late 2000s, seminal studies on IV discovered that *the negative power law*, a stylized fact regarding the term structures of IV, can be modeled consistently by means of fractional Brownian motions (Alòs et al. [2007] and Fukasawa [2011]).[2] Subsequently, Gatheral et al. [2018] reported that roughness can be observed from the behavior of HVs calculated from high-frequency data for equity indices and bond futures.

These studies heralded one of the recent trends in volatility research that utilizes fractional Brownian motions to express roughness.[3] In particular, volatility models using fractional

[1]The standard deviation of the change in a Brownian motion over a time length Δ is $\Delta^{1/2}$, whereas the corresponding standard deviation of a fractional Brownian motion is Δ^H, where $H \in (0,1)$ is the Hurst parameter representing the degree of roughness. When $H < 1/2$, the standard deviation of the change in a fractional Brownian motion is large for short time intervals (i.e., small Δ) compared to a Brownian motion, that is, a fractional Brownian motion fluctuates more volatilely than a Brownian motion. Note that roughness is generally defined based on the Hölder continuity of sample paths (see section 2.2).

[2]The negative power law is a stylized feature related to the term structure of IV (i.e., the relationship between IVs calculated from derivatives of different maturities) and is observed mainly in the U.S. equity derivatives market, where volatility trading is prevalent. For details, see section 4.2.

[3]One of the authors of Gatheral et al. [2018], Jim Gatheral (State University of New York), was awarded *Risk Magazine*'s Quant of the Year 2021 for his contributions to rough volatility research. This indicates that his work has attracted a great deal of attention from both researchers and practitioners. He worked at financial

Brownian motion can represent the behaviors of HV and IV mentioned above, and there have been many applied studies, such as ones on pricing of derivatives including volatility derivatives (e.g., options whose underlying asset is VIX). In addition, some studies have analyzed rough volatility from the market microstructure (MM) perspective and shed light on what kind of order placing behaviors yield roughness in the volatility. In this way, research on rough volatility has been expanding its coverage from studies on its measurement and mathematical properties to applied studies on derivatives practices and interdisciplinary studies on economic rationales for why volatility is rough.

Despite these developments, there are still a number of issues that require further research. First, discussions over whether volatility is rough or not remain ongoing. On the one hand, as mentioned above, many studies suggest the existence of roughness and potential applications of rough volatility. On the other hand, some studies point out the possibility of bias in the estimated values of roughness indices that may result in spurious detections of roughness. There are also ongoing discussions about whether fractional Brownian motion is really an appropriate ingredient to model roughness.[4] In addition, further technological and theoretical developments toward efficient numerical computation of rough volatility models are required for practical applications.

Against the above backdrop, this paper surveys cutting-edge research on rough volatility alongside with the history of existing volatility research from the perspectives of HV, IV, and MM. Research on HV roughly corresponds to financial econometrics, which focuses mainly on the time-series analysis of asset price fluctuations, while research on IV roughly corresponds to mathematical finance, which focuses mainly on derivative pricing. MM is a subset of financial economics, which mainly studies asset price formation mechanisms from an economic point of view. We first summarize the history of each field and its relation to practical issues in the financial industry, as well as that to academic issues in the economics, and then we discuss the position of rough volatility research within each research field and its prospects therein. Through these discussions, we describe the possibility and necessity of cross-disciplinary research with volatility as a nexus, as well as the prospects for practical applications of rough volatility.

First, section 2 introduces mathematical definitions and results for later use. Specifically, we introduce Brownian motion and Itô calculus, which have been fundamental tools in the study of volatility, especially in mathematical finance. Then, we introduce the definition and properties of fractional Brownian motion, focusing on its similarities and differences with conventional Brownian motion. We also provide an intuitive interpretation of roughness.

Section 3 focuses on HV. After reviewing the history of volatility research in the financial econometrics, we summarize Gatheral et al. [2018], a seminal study on the roughness of HV, and subsequent studies. Research on HV has been developed with the goal of constructing

---

institutions before becoming a professor at a university, suggesting that practical issues motivated his research as well.

[4]Currently, a majority of the rough volatility literature uses fractional Brownian motion, but this does not mean that it is the only modeling tool for roughness; that other stochastic processes can express observed rough behaviors has not been ruled out.

models that better represent the observed time-series features of HV and have better volatility forecasting accuracy. In particular, a vast literature has developed models that express the leverage effect and the long-memory properties of HV. It is also important to note that the analytical methods of HV have evolved along with increases in the availability of asset price data. In particular, since the late 1990s, the *realized volatility*, which is measured based on high-frequency asset price data (e.g., intraday data), has become the main variable of interest in HV research.

In this context, Gatheral et al. [2018], whose working paper version appeared in 2014, is one of the key papers that have raised the profile of rough volatility. Gatheral et al. [2018] showed that the time series of realized volatility exhibits roughness and reported that a time-series model that takes into account roughness could improve the accuracy of volatility forecasting. We distill two main issues by surveying the subsequent studies. The first issue is the estimation of roughness (typically measured as the Hurst parameter of fractional Brownian motion). The realized volatility is subject to measurement errors, as it is an approximation of volatility described by a continuous-time model. Therefore, some studies propose sophisticated estimation procedures for roughness that are robust to measurement errors in the realized volatility. The second issue is about the relationship between roughness and other traditionally considered features of volatility, especially the long-memory property. The long-memory property is the fact that volatility is a highly persistent process that exhibits positive autocorrelation over a long period of time. Since roughness suggests extreme fluctuations in a short period, further theoretical and empirical investigations of the relationship between these two properties are called for.

Section 4 focuses on IV. After reviewing the history of mathematical finance and related practical issues in the derivatives business, we provide an overview of rough volatility research in this area. Derivative pricing models are used for pricing complex derivative products and risk management (i.e., derivation of hedging strategies) after calibrating (adjusting) model parameters to represent market prices of highly liquid derivatives. For this reason, it is a prerequisite for derivative pricing models to have the ability to fit to the market prices of highly liquid financial instruments. Since IV is calculated from derivative prices, it is necessary to construct a model that can represent IV. Therefore, more sophisticated derivative models that can represent the IV observed in the derivatives market have been developed, especially since the 1990s.

In this context, in the early 2000s, a negative power law for the term structure of IV was observed in the U.S. equity derivatives market, and it was pointed out that it is difficult to reconcile this observation with conventional models based on Brownian motion (e.g., Carr and Wu [2003]). In the late 2000s, Alòs et al. [2007], and later in Fukasawa [2011], showed that a rough volatility model based on fractional Brownian motion could represent the negative power law of IV. These studies are considered to be the pioneers of rough volatility research. In recent years, research on this issue has progressed, and some studies have provided strong support for the need for rough volatility. Specifically, Fukasawa [2021] proved that a rough volatility model with an appropriate roughness parameter (i.e., the roughness consistent with the one implied by the negative power law of IV) is necessary for pricing and hedging derivatives, because otherwise inconsistencies (i.e., arbitrage opportunities) would arise. In addition to these theoretical studies, studies on practical applications have been conducted. Unlike conventional Brownian motion, fractional Brownian motion does not have nice properties such as the Markov property (i.e., future behaviors do not depend on past behaviors), so it has been considered difficult to apply conventional mathematical finance theory to fractional Brownian motion. Related to this, recent research has revealed that the Markov property "recovers" under certain conditions (Bayer et al. [2016]). This demonstrates how such mathematical difficulties are gradually being overcome, increasing the applicability of rough volatility models to derivatives practices. On the other hand, it is noteworthy that, at present, the derivatives markets in which roughness is observed (i.e., markets in which the negative power law is observed) are limited to a few markets, such as the U.S. equity derivatives market. Therefore, it is still an open question whether roughness should be taken into account in pricing and risk management for other markets (e.g., other countries and other asset classes). In addition, the practical application of rough volatility models requires much more complex numerical calculations than those based on conventional Brownian motion and Itô calculus, so further development of theory and numerical calculation techniques is needed.

Section 5 focuses on MM. We briefly overview the literature on MM so as to show the relation between rough volatility research and financial economics. The MM literature studies, as its name suggests, how micro-level features, such as the institutional structure of exchanges and the behaviors of individual market participants, impact asset price formation. A subset of MM studies are aimed at deriving stochastic processes of asset price dynamics, starting from assumptions on *micro-foundations* such as the buy and sell order placing behaviors of individual market participants. In rough volatility research, some studies have investigated the conditions under which market participants' buy and sell order behaviors lead to roughness in the volatility of the asset prices. For example, El Euch et al. [2018] showed that volatility exhibits rough behavior when price updates from a particular period of trading continue to affect the frequency of future price update occurrences with a certain degree of persistence. Such studies are considered attempts to identify the source of roughness from a financial economic viewpoint, which is an important issue from the perspective of the institutional design of markets and its relation to financial stability. However, since research to date has presented only a mathematical framework, it is necessary to develop further research on more realistic modeling of investor behaviors and empirical analysis using actual data.

In summary, this paper is a cross-disciplinary survey of the fields of financial econometrics, mathematical finance, and financial economics, with volatility research as a nodal point. Therefore, this paper can be read in various manners, depending on readers' knowledge and interest. Readers who wish to get an overview of the study can skip section 2, since it summarizes mathematical results necessary to understand the technical aspects in the later sections. In addition, readers can read sections 3, 4, and 5 independently, in any order, although these sections are interrelated.

2 Mathematical preparation: fractional Brownian motion

In this section, we prepare mathematical definitions and results for their later use. We first introduce Brownian motion and Itô calculus (stochastic differential equations) in section 2.1, and then describe the formulation and properties of fractional Brownian motion in section 2.2. Although these concepts are usually formulated based on measure-theoretic probability theory, we omit rigorous mathematical descriptions and explain only the necessary facts as intuitively as possible. For details, see Karatzas and Shreve [1998] for a mathematical treatment of Brownian motion and Itô calculus, and Nualart [2006] for fractional Brownian motion.

Note that, as discussed in sections 3 and 4, in rough volatility studies, fractional Brownian motion is often used to model the logarithm of volatility.[5]

2.1 Brownian motion and Itô calculus

2.1.1 Definition and properties of Brownian motion

A mathematical model of a variable that randomly fluctuates along with time is called a stochastic process. Mathematically speaking, a collection of random variables $\{X_t \mid t \geq 0\}$ (parameterized by a time parameter t) defined on a probability space $(\Omega, \mathcal{F}, P)$ is called a stochastic process. The value this stochastic process takes at time t can be viewed as a random variable, which we denote by X_t. If we fix a sample point $\omega \in \Omega$ and consider $X_t(\omega)$ as a function with respect to time, we call it a sample path of the stochastic process for a sample point ω.

Brownian motion is one of the most fundamental continuous-time stochastic processes.[6] A Brownian motion is the continuous limit of a random walk in a discrete-time model and has various important properties, such as the fact that its increments (the time variation $X_{t+\Delta} - X_t$) follow independent normal distributions. Formally, Brownian motion is defined as a stochastic process $\{W_t \mid t \geq 0\}$ satisfying the following properties.[7]

[5]For instance, the instantaneous volatility σ_t is modeled as $\sigma_t = e^{F(B_t^H)}$, where B_t^H is a fractional Brownian motion and F is some function. See section 3.2 for more details.

[6]Since the beginning of the 20th century, the mathematical formulation and properties of Brownian motion have been studied, mainly with a focus on modeling various phenomena in physics and economics. In particular, the relationship between Brownian motion and the field of finance goes back to around 1900. Prior to extensive use in mathematics and physics, Bachelier [1900] discussed option pricing using a model that could be considered a prototype of Brownian motion. Subsequently, in the early 1900s, Kolmogorov, Wiener, Kiyoshi Itô, and others developed theories such as measure-theoretic probability theory and Itô calculus. As is well known, Merton [1973] presented an option pricing method based on Itô calculus, which heralded developments in the fields of financial engineering and mathematical finance.

[7]Note that the existence of stochastic processes satisfying the given conditions is not trivial. For a proof of the existence of Brownian motion, see, for example, section 2 of Karatzas and Shreve [1998].

$$
\begin{cases}
\text{(i) Zero initial value :} \\
\qquad\qquad P(\omega \mid W_0 = 0) = 1. \\
\text{(ii) Continuty of sample path：} \\
\qquad\qquad P(\omega \ \mid \lim_{s \to t} W_s(\omega) = W_t(\omega)) = 1. \\
\text{(iii) Increments follow a normal distribution (mean: 0, variance: time-increment)：} \\
\qquad\qquad W_t - W_s \sim N(0, t-s), \ \text{for } 0 \le s < t. \\
\text{(iv) Increments are independent：For } 0 \le t_0 < t_1 < \cdots < t_k < \infty, \\
\qquad\qquad W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \ldots, W_{t_k} - W_{t_{k-1}} \text{ are indepenedent.}
\end{cases}
\tag{1}
$$

It is well known that Brownian motion has important properties such as the Markov property and martingale property. The Markov property is that the conditional probability distribution of a stochastic process at a future time depends only on the information at the present time and not on the past history (information accumulated prior to the present time). For example, given the value W_s of a Brownian motion at time s, the conditional probability distribution of the Brownian motion at a future time t $(> s)$ satisfies the relation $P\left(W_t \in B \mid \mathcal{F}_s\right) = P(W_t \in B \mid W_s)$. Here, both sides are conditional probabilities that W_t is in some subset $B \subset \mathbb{R}$. The left-hand side is a probability distribution conditioned on the information of Brownian motion up to time s, and the right-hand side is a probability distribution conditioned only on the information that the value of Brownian motion at time s equals W_s. The martingale property is the so-called fair betting game property, that is, the conditional expected value of a stochastic process at a future point in time equals the current value. Mathematically, this property is expressed as $E[W_t \mid \mathcal{F}_s] = W_s$. These two properties of Brownian motion are derived from the property that a Brownian motion is an independent incremental process with mean zero.

A Brownian motion has a bounded and nonzero quadratic variation defined up to time T (> 0) as follows:

$$\langle W \rangle_{2,T} = \lim_{n \to \infty} \sum_{i=0}^{n-1} \left(W_{\frac{i+1}{n}T} - W_{\frac{i}{n}T} \right)^2 .$$

As shown above, quadratic variation is expressed in terms of sums of squared increments of a stochastic process for increments representing finely divided steps within a certain period of time. Quadratic variation is known to be zero for smooth processes, such as differentiable sample paths, whereas a Brownian motion, which randomly fluctuates and has non-smooth sample path, has a positive quadratic variation. Specifically, the above equation converges to T with probability one; that is,

$$P\left(\langle W \rangle_{2,T} = T\right) = 1$$

holds. This follows from the definition of Brownian motion (1)(iii), $\mathrm{Var}\left[W_{t_i} - W_{t_j}\right] = t_i - t_j$.

2.1.2 Itô calculus

One of the most important features of Brownian motion is that one can define integration by Brownian motion, as what is known as an *Itô integral*. To define an Itô integral with respect to an integrand $h(s)$, we proceed as follows. For a time division $0 = s_0 < s_1 < \cdots < s_n = t$, we multiply the increment of Brownian motion $W_{s_{i+1}} - W_{s_i}$ by $h(s_i)$, then add up all increments (note that the value of $h(s)$ is the value at the starting time of the increment of Brownian motion):

$$\sum_{i=0}^{n-1} h(s_i)\left(W_{s_{i+1}} - W_{s_i}\right).$$

Then, we can define the limit of this summation as the time division becomes finer in a mathematically rigorous manner. Specifically, we can define this limit as the Itô integral $\int_0^t h(s)\, dW_s$, that is,

$$\int_0^t h(s)\, dW_s = \lim_{n\to\infty} \sum_{i=0}^{n-1} h(s_i)\left(W_{s_{i+1}} - W_{s_i}\right).$$

Note that the rigorous formulation of this limit and convergence is the core of Itô calculus, but we omit the mathematical details here since it requires additional preparations for an accurate description.[8]

A stochastic process that adds an ordinary integral term to an Itô integral is called an Itô process. Namely, an Itô process X_t is a stochastic process expressed as follows:

$$X_t = x + \int_0^t \mu(s, X_s)\ ds + \int_0^t \sigma(s, X_s)\ dW_s. \tag{3}$$

Here, the function μ represents the trend and the function σ represents the magnitude of the fluctuation of X_s. In equation (3), the second and third terms are called the drift and volatility

[8]The following is a sketch of this point. First, when the integrand h is an indicator process (i.e., a piecewise constant stochastic process $h(s) = \sum_{i=0}^{n-1} a_i \mathbb{1}_{\{s_i \le s < s_{i+1}\}}$), the Itô integral can be naturally defined as $\int_0^t h(s) dW_s = \sum_{i=0}^{n-1} h(s_i)\left(W_{s_{i+1}} - W_{s_i}\right)$. For a general stochastic process h, we rely on the fact that h can be approximated by a sequence of indicator processes. Considering a sequence of Itô integrals for each sequence of indicator processes approximating h, we can show that this sequence also converges. This limit is defined as the Itô integral for the more general stochastic process h.

When discussing such approximation and convergence, it is necessary to define an appropriate functional space and a norm (distance) on that space. Specifically, for the functional space of integrands, we consider the square integrable stochastic process space $L^2 = \left\{h(s) \mid E\left[\int_0^t h^2(s)\, ds\right] < \infty\right\}$ equipped with the inner product $\langle h, g\rangle = E\left[\int_0^t h(s)g(s)ds\right],\ h, g \in L^2$. From the definitions (iii) and (iv) of Brownian motion, we can see that $E\left[\left(\sum_{i=0}^{n-1} h(s_i)\left(W_{s_{i+1}} - W_{s_i}\right)\right)\left(\sum_{i=0}^{n-1} g(s_i)\left(W_{s_{i+1}} - W_{s_i}\right)\right)\right] = \sum_{i=0}^{n-1} h(s_i)\, g(s_i)\,(s_{i+1} - s_i)$. This result proves that the destination space of the Itô integral (viewed as a mapping operator) is also a Hilbert space, with the inner product

$$\left\langle \int_0^t h(s)dW_s, \int_0^t g(s)dW_s \right\rangle = \langle h, g\rangle = E\left[\int_0^t h(s)g(s)ds\right]. \tag{2}$$

Thus, the above-mentioned sequence of indicator processes has a limit. Note that equation (2) indicates that a formal calculation $dW_t \cdot dW_t = dt$ is valid, corresponding to the fact that the quadratic variation of a Brownian motion is T, that is, $P\left(\langle W\rangle_{2,T} = T\right) = 1$. This formula is the celebrated Itô lemma.

terms, respectively. Equation (3) can also be written in differential form

$$dX_s = \mu(s, X_s)\ ds + \sigma(s, X_s)\ dW_s, \quad X_0 = x,$$

which is called a stochastic differential equation. Stochastic differential equations make it possible to describe more complex models based on Brownian motion (e.g., models in which drift or the volatility changes). Note that, like a Brownian motion, Itô process X_t is a Markov process. It is also closely related to the martingale property.[9]

2.2 Fractional Brownian motion

As will be discussed in section 3, in volatility modeling, there are a number of observations that cannot be represented by conventional Brownian motion and Itô processes. Therefore, other models that can represent these stylized facts have been developed. Typical examples of volatility fluctuations that cannot be represented by Brownian motion and Itô processes are as follows.

1. While increments of Brownian motion are independent, actual time series of volatility are autocorrelated.
2. While variances of Brownian motion are proportional to time increments, actual time series of volatility do not satisfy this property.[10]

Fractional Brownian motion, which is an extension of Brownian motion, is known for a model that can describe these properties, and has been used to describe the stochastic dynamics of volatility.[11] Fractional Brownian motion has the property that its increments follow a normal distribution, which is the same as Brownian motion. On the other hand, unlike a Brownian motion, it is a stochastic process that has a parameter called the Hurst parameter and may exhibit nonzero autocorrelation (past and future increments are not independent).

Fractional Brownian motion $B^H = \left\{B_t^H \mid t \geq 0\right\}$ with the Hurst parameter $H \in (0, 1)$ is defined as a stochastic process that satisfies the following conditions.

[9] Itô integral $Y_t = \int_0^t \sigma(s)\ dW_s$ has the martingale property. Conversely, it is known that any continuous stochastic process with the martingale property can be represented as an Itô integral. As is well known, the martingale property plays a central role in the formulation of derivative pricing principles such as no-arbitrage and replication. The Markov property is also an extremely important property for numerical computation of derivative prices and of associated risk amounts in practical applications.

[10] In the case of models based on Brownian motion, the variance of daily data is one-fifth of the variance of weekly data. On the other hand, empirical studies have revealed that high-frequency data are more volatile than what Brownian motion suggests.

[11] Note that Brownian motion and Itô processes are still major modeling tools for the underlying asset price process.

$$
\begin{cases}
\text{(i) Zero initial value: } P\left(\omega \mid B_0^H = 0\right) = 1. \\
\text{(ii) Continuity of sample paths: } P\left(\omega \;\middle|\; \lim_{s\to t} B_s^H(\omega) = B_t^H(\omega)\right) = 1. \\
\text{(iii) Increments follows a normal distribution:} \\
\qquad\qquad B_t^H - B_s^H \sim N\left(0, (t-s)^{2H}\right), \text{ for } 0 < s < t. \\
\text{(iv) Gaussian increments: For } 0 \le t_0 < t_1 < \cdots < t_k < \infty, \\
\text{increments } B_{t_1}^H - B_{t_0}^H, B_{t_2}^H - B_{t_1}^H, \ldots, B_{t_k}^H - B_{t_{k-1}}^H \text{ follow a joint normal distribution.}
\end{cases}
\tag{4}
$$

Namely, the definitions of Brownian motion (1)(iii)–(iv) are replaced by (4)(iii)–(iv). When $H = 1/2$, fractional Brownian motion boils down to Brownian motion, that is, $W_t = B_t^{1/2}$.

Unlike Brownian motion, when $H \neq 1/2$, the increments of fractional Brownian motion are not independent. In particular, one of the following relationships holds, depending on the value of the Hurst parameter.[12] For $s_1 < t_1 < s_2 < t_2$,

$$
\begin{cases}
H \in \left(0, \frac{1}{2}\right) & \Rightarrow \quad E\left[\left(B_{t_1}^H - B_{s_1}^H\right)\left(B_{t_2}^H - B_{s_2}^H\right)\right] < 0, \\
H = \frac{1}{2} & \Rightarrow \quad E\left[\left(B_{t_1}^H - B_{s_1}^H\right)\left(B_{t_2}^H - B_{s_2}^H\right)\right] = 0, \\
H \in \left(\frac{1}{2}, 0\right) & \Rightarrow \quad E\left[\left(B_{t_1}^H - B_{s_1}^H\right)\left(B_{t_2}^H - B_{s_2}^H\right)\right] > 0.
\end{cases}
$$

This fact indicates that the increments are negatively correlated in the case of $H < 1/2$ and positively correlated in the case of $H > 1/2$. It can be seen that fractional Brownian motion ($H \neq 1/2$) is neither a martingale and nor a Markov process. It is also known that fractional Brownian motion has the properties of self-similarity (for $a > 0$, $B_{at}^H \sim a^H B_t^H$) and time homogeneity ($B_{t+s}^H - B_s^H \sim B_t^H$).

Furthermore, from definitions (4)(i) and (iii), the following equation holds:

$$
\sqrt{\mathrm{Var}\left[(B_t^H)^2\right]} = t^H. \tag{5}
$$

Namely, the time evolution of the standard deviation is time raised to the power of H. This is a generalization of the well-known *square-root-of-time rule* for Brownian motion, which states that the standard deviation of a Brownian motion increases proportional to the square

---

[12] From (4)(iii), for $s, t > 0$, $E\left[B_t^H B_s^H\right] = \frac{1}{2} E\left[\left(B_t^H\right)^2 + \left(B_s^H\right)^2 - \left(B_t^H - B_s^H\right)^2\right] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right)$ holds. Then we have

$$
E\left[\left(B_{t_1}^H - B_{s_1}^H\right)\left(B_{t_2}^H - B_{s_2}^H\right)\right] = \frac{1}{2}\left((t_2 - s_1)^{2H} - (t_2 - t_1)^{2H} - (s_2 - s_1)^{2H} + (s_2 - t_1)^{2H}\right).
$$

If we let $a_1 = t_2 - s_1, a_2 = t_2 - t_1, b_1 = s_2 - t_1, b_2 = s_2 - s_1$, then we see that $a_1 + b_1 = a_2 + b_2 = (t_2 - t_1) + (s_2 - s_1) > 0$, and $a_1 > a_2, b_2 > b_1$. Setting $f(x) = x^{2H}$, f is a concave function in the case of $H < 1/2$ and a convex function in the case of $H > 1/2$; thus, the following holds:

$$
E\left[\left(B_{t_1}^H - B_{s_1}^H\right)\left(B_{t_2}^H - B_{s_2}^H\right)\right] = \frac{f(a_1) + f(b_1)}{2} - \frac{f(a_2) + f(b_2)}{2}
\begin{cases}
< 0 & (H < 1/2), \\
= 0 & (H = 1/2), \\
> 0 & (H > 1/2).
\end{cases}
$$

root of time t.

This fact implies that a fractional Brownian motion with Hurst parameter $H < 1/2$ is a stochastic process that is highly volatile in a short time (a rough stochastic process). In other words, while the value at some time t follows a normal distribution, the same as a Brownian motion, the standard deviation of the normal distribution at an early time point (when t is small) is larger than that of a Brownian motion (Figure 1).

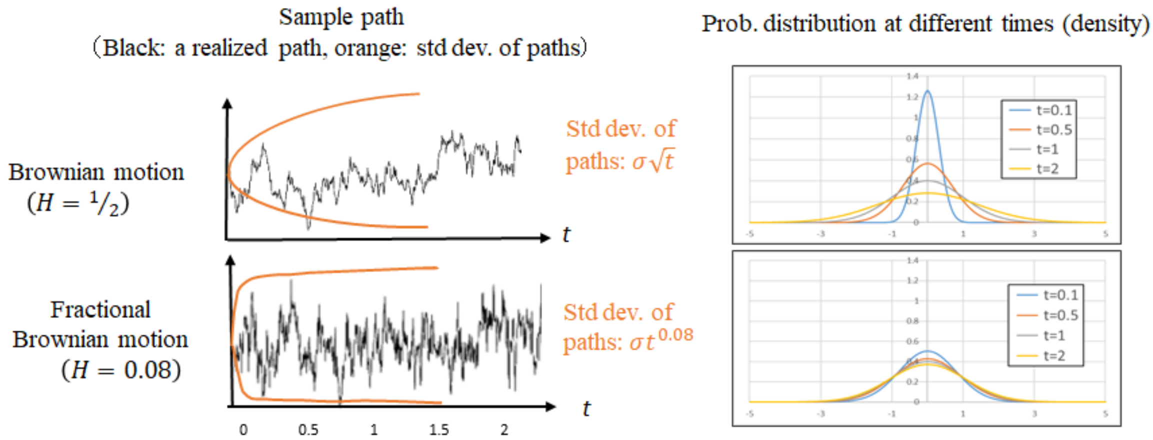

Figure 1: Sample paths and their dispersion of fractional Brownian motion

For fractional Brownian motion, when $H < 1/2$, the short-time fluctuations are more intense than Brownian motion, that is, the sample path exhibits rough behavior (Figure 2, upper left), whereas when $H > 1/2$, the fluctuations are milder, that is, the sample path exhibits smooth behavior (Figure 2, upper right) relative to Brownian motion.

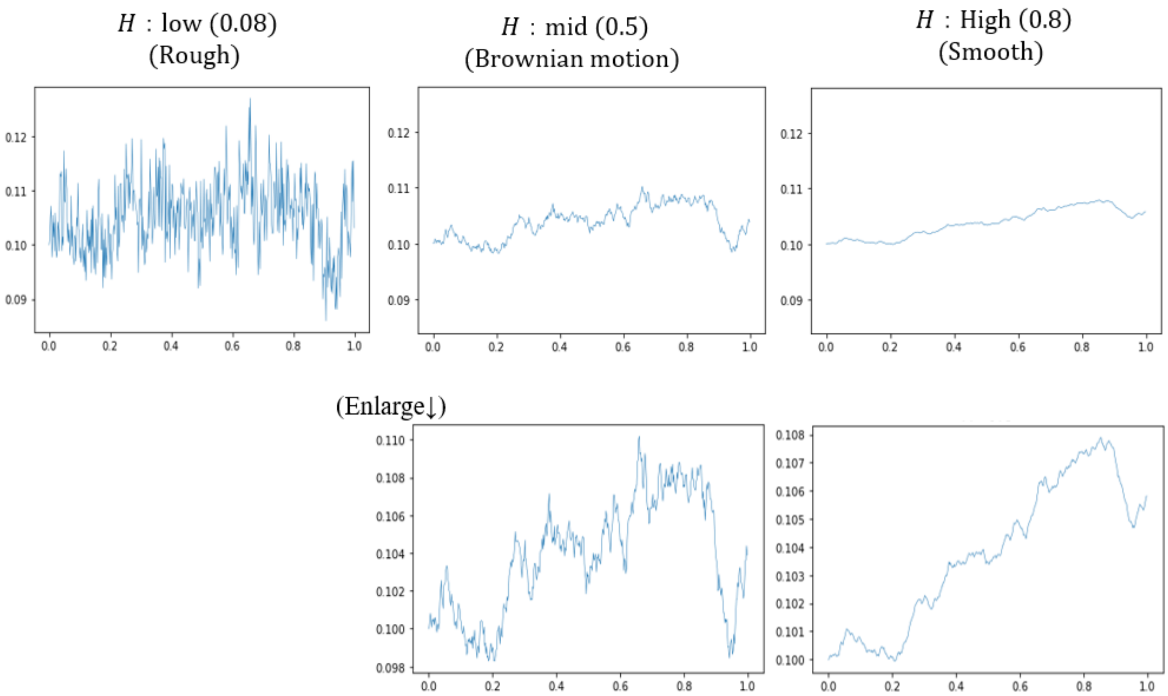

Figure 2: Sample paths of fractional Brownian motion with different H

Note that roughness is generally defined as the Hölder continuity of *paths* (i.e., functions parameterized by time). The Hölder continuity describes the regularity of the sample path (how extremely the path fluctuates), from which the term *rough* is derived. Since it is known that the fractional Brownian motion of the Hurst parameter H is $(H-\varepsilon)$−Hölder continuous, the Hurst index can be viewed as an index of the Hölder continuity of the fractional Brownian motion.[13]

Similarly, while the quadratic variation is finite in the Brownian motion case, equation (5) indicates that, in the fractional Brownian motion case, $1/H$-variation is finite and the following holds:[14]

$$\langle B \rangle_{1/H,T} = \lim_{n\to\infty} \sum_{i=1}^{n} \left(B^H_{\frac{i}{n}T} - B^H_{\frac{i-1}{n}T} \right)^{1/H}, \quad P\left(\langle B \rangle_{1/H,T} = T \right) = 1.$$

In addition, fractional Brownian motion is represented as an Itô integral with an integrand

[13] A process $\{B_t^H \mid t \geq 0\}$ is said to be $(H-\varepsilon)$−Hölder continuous if there exists a continuous modification $\{\tilde{B}_t^H \mid t \geq 0\}$ such that for any $\varepsilon > 0$, there exists a random variable G_ε such that $\tilde{B}_t^H - \tilde{B}_s^H \leq G_\varepsilon |t-s|^{H-\varepsilon}$ holds.

[14] Therefore, if we attempt to define the integral with respect to fractional Brownian motion following the same approach as the Itô integral, we cannot apply the formal calculation corresponding to the Itô lemma ($dW_t \cdot dW_t = dt$). This means that while the formulation of the Itô integral relies on symmetry and inner product structure in the Hilbert space in order to define the integral, in the case of fractional Brownian motion, it is necessary to discuss a Banach space which has neither symmetry nor inner product structure. Another approach to formulate integrals with respect to fractional Brownian motion has recently been developed in rough path theory (Lyons [1998]). For example, Bayer et al. [2020] discussed analysis of rough volatility models based on regularity structure theory, which won the 2014 Fields Prize, and further applications to rough volatility are expected in the future.

that has poles as follows (Levy [1953], Mandelbrot and Van Ness [1968]):

$$B_t^H = \frac{1}{C_H}\left(\int_0^t (t-s)^{H-\frac{1}{2}}\, dW_s + \int_{-\infty}^0 \left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right) dW_s\right), \tag{6}$$

where $C_H = \left(\int_0^\infty \left((1+s)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}}\right)^2 ds + \frac{1}{2H}\right)^{\frac{1}{2}}$. From equation (6), we can calculate the time-series correlation of fractional Brownian motion (4)(iii) by considering the Itô integral including the past history of Brownian motion, as follows:

$$\begin{aligned} E\left[\left(B_t^H - B_s^H\right)^2\right] &= \frac{1}{C_H^2} E\left[\left(\int_{-\infty}^{\infty} (\max\{t-u,0\})^{H-\frac{1}{2}} - (\max\{s-u,0\})^{H-\frac{1}{2}}\ dW_u\right)^2\right] \\ &= \frac{1}{C_H^2} E\left[\int_{-\infty}^{\infty} \left((\max\{t-u,0\})^{H-\frac{1}{2}} - (\max\{s-u,0\})^{H-\frac{1}{2}}\right)^2\ du\right] \\ &= (t-s)^{2H}. \end{aligned}$$

We can see that, the first term on the right-hand side of equation (6), $\tilde{B}_t^H = \int_0^t (t-s)^{H-1/2}\, dW_s$, is a continuously weighted sum of a Brownian motion. Note that $\tilde{B}_t^H$ is called the Riemann–Liouville-type Itô integral[15] and is often used in rough volatility modeling.

3 Observation of roughness in historical volatility

In this section, first we summarize the importance of volatility research in section 3.1. Then, by focusing on HV, we provide an overview of the development of volatility research in the field of financial econometrics.[16] Next, section 3.2 provides a summary of Gatheral et al. [2018], which is one of the seminal articles in rough volatility research. This article reports empirical evidence of the *roughness* of the time series of realized volatility observed in high-frequency data for various financial assets. In section 3.3, we overview studies subsequent to Gatheral et al. [2018] on rough volatility research.

3.1 History of HV research in financial econometrics

3.1.1 Importance of volatility research

In the field of finance, major research topics include the analysis of optimal investment behavior and asset price formation mechanisms under the presence of uncertainty. For example, in the mean-variance portfolio theory of Markowitz [1952], the investment weights of

[15] The Riemann–Liouville-type integral is defined for the usual Riemann integral. These studies have their origin in the era of Leibniz (15th to 16th century). For a function f, the Riemann–Liouville-type integral of order λ is defined as

$$I^\lambda f = \frac{1}{\Gamma(\lambda)} \int_0^t (t-s)^{\lambda-1} f(s)\, ds,$$

which is an extension of the Reimann integral to the fractional order. This is the origin of the term *fractional*.

[16] For an overview of volatility research in financial econometrics, see Takahashi et al. [2023].

an optimal portfolio are given as a function of the expected returns and the covariance matrix of individual stocks. In the Capital Asset Pricing Model (CAPM), one of the prominent asset pricing models, the expected returns of individual stocks are explained by the covariance between the return of the stocks and that of the market portfolio under the assumption of a representative mean-variance investor (Sharpe [1966], Lintner [1965], Mossin [1966]).

As such, moments of the probability distribution of asset returns play an important role in finance research. Accordingly, time-series analysis and modeling of moments of asset price fluctuations have been a central research theme, especially in financial econometrics. Among these studies, volatility (standard deviation or variance), the second moment of asset returns, has been attracting more attention than expected return (i.e., the first moment).

There are several reasons why research on volatility has progressed further than research on expected returns. The first is that volatility can generally be more accurately estimated thanks to advances in data availability and the development of quantitative methods, whereas expected return remains notoriously difficult to estimate precisely. For example, Merton [1980] showed that under the assumption that there are no measurement errors in observed asset prices, volatility can be more accurately estimated as higher frequency data become available. Contrarily, he pointed out that accurate estimation of expected return is difficult even when high-frequency data are available. In addition, expected returns are known to exhibit large stochastic variations, and estimation from historical data is fraught with difficulties (e.g., Martin [2017]). With such theoretical background, vast number of studies have focused on volatility, which is relatively easy to estimate and forecast.[17]

The second reason is the importance of volatility in risk management. Knowledge on the magnitude of price fluctuations during sudden market turmoil is essential in measuring and analyzing market risks. This implies the importance of the measurement of volatility (and higher-order moments, which capture the shape of the probability distribution of asset returns) in risk management, as volatility represents the magnitude of asset price fluctuations. Moreover, in the fields of mathematical finance and financial engineering, volatility is a key modeling target for derivative pricing and risk management purposes, whereas estimates of expected returns are not required. For more details, see section 4.

Volatility and uncertainty in general are also proven as important elements in macroeconomics; a large body of literature has shown that uncertainty has a negative impact on real economic activity (e.g., Bloom et al. [2007]). For this reason, the VIX index and the Nikkei Stock Average Volatility Index, which are volatility indicators calculated from option prices, are widely used in macroeconomic analysis as proxies of uncertainty in financial markets and the macroeconomy, and various studies have proposed alternative indicators of uncertainty (e.g., Bloom [2009]; Baker et al. [2016]; Dew-Becker et al. [2021]).

[17]In fact, a majority of empirical studies on optimal portfolio choices focus on minimum variance portfolios which use only the information on the covariance matrix, rather than mean-variance portfolios (e.g., Michaud [1989], Best and Grauer [1991], Jagannathan and Ma [2003]). This is because the empirical disadvantage of constructing mean-variance portfolios using estimates of expected returns, which may be contaminated by large estimation errors, greatly outweighs the theoretical advantage of incorporating the information on expected returns to construct a mean-variance portfolio.

As discussed above, volatility is one of the crucial variables in the various fields of finance and economics. For this reason, in the field of financial econometrics, vast amounts of effort have been devoted to analyzing the time-series features of HVs measured from historical data for actually observed asset prices and developing discrete-time volatility models that have better fit to the data and better forecast accuracy. The typical features of volatility fluctuations pointed out by many studies on HVs include *persistence of volatility (volatility clustering)*, *leverage effect*, and *long memory*. In the following, we overview the history of research on these features.

3.1.2 Persistence and leverage effect of HV

It has long been known that asset returns fluctuate randomly as their autocorrelation is almost zero. On the other hand, the time series of the absolute values of returns and the squared returns exhibit very long positive autocorrelations (e.g., Mandelbrot [1963], Fama [1965], Ding et al. [1993]). Since the absolute value of returns and the squared returns represent the magnitude of the variability of asset prices, this fact indicates that volatility exhibits strong autocorrelation and persistence. This phenomenon is known as *volatility clustering*. Regarding time-series models including persistence of volatility, autoregressive models of the variance of financial assets have been developed since the 1980s. Engle [1982] proposed the ARCH (autoregressive conditional heteroskedasticity) model, in which volatility is represented by an autoregressive model of squared noise of past asset price changes.[18] Bollerslev [1986] extended the ARCH model as a generalized autoregressive conditional heteroskedasticity (GARCH) model, in which volatility depends on autoregressive lag in addition to the ARCH model terms. These time-series models have been fundamental tools in the analysis of HV time series, as they are parsimonious yet achieve a good fit to the observed data and good forecasting accuracy.

A prominent and important feature, along with the persistence of volatility, is the negative correlation between the return of an equity index and its volatility. This feature means that volatility tends to increase as stock prices fall, which is commonly referred to as the *leverage effect*.[19] Under the presence of the leverage effect, the probability distribution of asset prices has a fat tail in the downside direction. Therefore, it is important to use a model that can appropriately represent the leverage effect from the perspective of risk management and portfolio selection. Models that extend the GARCH-type model to represent the leverage effect include the exponential GARCH model of Nelson [1991] and the GJR model of Glosten et al. [1993].

[18] For an overview of ARCH-type volatility models, see, for example, Chapter 21 of Hamilton [1994] and Poon and Granger [2003].

[19] The term leverage effect comes from the idea that when the stock price falls, the leverage of the firm, and thus the riskiness of the stock, and the volatility of stock price increases. See Black [1976] and Christie [1982]. Note that, some empirical studies have shown other factors contributing to the negative correlation between returns and volatility (Wu [2001], Avramov et al. [2006]).

3.1.3 Long memory

Among the properties of volatility fluctuations, a vast number of financial econometric studies have noted that asset price dynamics exhibit *long memory*.[20] For example, the aforementioned Ding et al. [1993] reported that time series of absolute values of daily returns exhibit significant positive autocorrelation even for extremely long lags, such as a 10-year lag. This feature implies that the autocorrelation function (which returns the autocorrelation coefficient as a function of the order of the lag) decreases slowly. On the other hand, the GARCH-type model described above has the limitation that it could not represent long memory, because the autocorrelation function decays exponentially. Therefore, research on volatility modeling based on time-series models with long memory has been actively pursued, especially since the 1990s.

A standard discrete-time model for representing long memory is the autoregressive fractionally integrated moving average (ARFIMA) model (Granger [1980], Granger and Joyeux [1980], Hosking [1981]). The ARFIMA model is an extension of the ARIMA model. For example, the ARIMA$(0, 1, 0)$ model, the simplest first-order integrated model, is equal to the noise term when the first-order difference is taken.[21] Namely, if L is the lag operator ($LX_t = X_{t-1}$), it can be represented as $(1 - L)X_t = \varepsilon_t$. The ARFIMA model extends this difference order to real numbers: the dth-order ARFIMA model ARFIMA$(0, d, 0)$ is represented as $(1 - L)^d X_t = \varepsilon_t$, where the left-hand side is formally defined as follows:

$$(1 - L)^d X_t = X_t - dX_{t-1} + \frac{d(d-1)}{2} X_{t-2} - \cdots .$$

The ARFIMA model is a stationary process when $-1/2 < d < 1/2$. When $0 < d < 1/2$, the autocorrelation function decays with the order of k^{2d-1}, where k is the lag order. The ARFIMA model also satisfies the following:

$$\lim_{n\to\infty} \sum_{j=-n}^{n} \left|\rho_j\right| = \infty.$$

This property implies a slow decay of the autocorrelation function, and is one definition of long memory. On the other hand, in the autoregressive moving average (ARMA) model and other models, the sum of the absolute values of the above autocorrelation functions is finite. Thus, when $0 < d < 1/2$, the ARFIMA model can be regarded as a time-series model that lies between a stationary $I(0)$ process without long memory (e.g., ARMA model) and a

[20] The existence of long memory has a considerable history in fields that including hydrology (e.g., the study of time series of water flow by Hurst [1951], which is the origin of the term Hurst index). In economics, Granger [1966] and Nerlove [1964] conducted spectral analysis of major macroeconomic variables and observed that they have a large low-frequency component (i.e., the contribution of the long-cycle component is large). Haubrich and Lo [1991] and Sowell [1992] analyzed macroeconomic variables based on the ARFIMA model, which is a time-series model with long memory mentioned in the main text. See a survey by Baillie [1996] for more details.

[21] In general, the ARIMA(p, d, q) model is a dth-order integrated model with pth-order autoregressive terms and qth-order moving average terms. The dth-order integrated model becomes a stationary process by differencing d times, and is called an $I(d)$ process.

non-stationary $I(1)$ process (e.g., random walk).

Being related to the subject of this paper, rough volatility and fractional Brownian motion, the relation between the ARFIMA model and FGN (fractional Gaussian noise) model is worth mentioning. The FGN model is a discrete-time model defined as the difference of fractional Brownian motion. The autocorrelation of the ARFIMA$(0, d, 0)$ process decays with a power exponent similar to the FGN whose Hurst parameter is $H = d + 1/2$. This relation demonstrates that ARFIMA$(0, d, 0)$ with $0 < d < 1/2$, a discrete-time time-series model with long memory, is closely related to fractional Brownian motion with Hurst parameter greater than $1/2$.

Early studies that took into account the long memory in the GARCH model include those of Baillie et al. [1996], who proposed the fractionally integrated generalized autoregressive conditional heteroskedasticity (FIGARCH) model, and Bollerslev and Mikkelsen [1996], who proposed the fractionally integrated exponential generalized autoregressive conditional heteroskedasticity (FIEGARCH) model. Breidt et al. [1998] proposed a discrete-time stochastic volatility model with long memory, and continuous-time models by Comte and Renault [1996], Comte and Renault [1998], Cheridito et al. [2003], Comte et al. [2012], and others have used fractional Brownian motion to express long memory. In these studies, to represent long memory, the discrete-time ARFIMA model with fractional integration parameter d such that $0 < d < 1/2$ and continuous-time fractional Brownian motion with a Hurst parameter H larger than $1/2$ are considered.

3.1.4 Realized volatility

In the early years of financial econometric research, when many of the studies mentioned above were conducted, asset price data with daily or longer frequency were mainly used due to data constraints. Since the late 1990s, the availability of high-frequency data for intraday asset price fluctuations has expanded. This expansion of data availability has promoted progress in financial econometrics in both theoretical and empirical aspects as described below.

When intraday price data are available, it is possible to calculate, say, the daily volatility from time-series data for five-minute price data within a single business day. The volatility calculated from such high-frequency price data is called the *realized volatility* and is a standard measure of HV in current financial econometrics. Specifically, realized volatility is calculated as follows:[22]

$$\mathrm{RV}\,(t, t+1, \pi) = \sum_{i=0}^{n-1} \left(\log S_{s_{i+1}} - \log S_{s_i}\right)^2,$$

where S_t is the underling price and π is a partition of $[t, t+1]$, $t = s_0 < s_1 < \cdots < s_n = t+1$.

[22]This definition of realized volatility (like integrated volatility, which will be defined later) corresponds to variance, not volatility (standard deviation), but following the convention in the field of financial econometrics, we refer to it as realized (integrated) volatility.

For example, realized volatility could be calculated from five-minute data (i.e., $s_{i+1} - s_i = 5$ min).

Andersen et al. [2001], Andersen et al. [2003], and Barndorff-Nielsen and Shephard [2002], among other studies, have revealed that the realized volatility defined in this way has theoretically favorable properties. Since this point is closely related to Gatheral et al. [2018], which will be discussed in more detail later, we give a brief explanation here.

In a continuous-time model, an underlying asset price with time-varying volatility is generally represented by the following stochastic differential equation (see section 2 for notation):

$$\frac{dS_t}{S_t} = \mu_t \, dt + \sigma_t \, dW_t,$$

where σ_t represents the instantaneous volatility at time t.[23] Given the aforementioned observations that volatility fluctuates stochastically, σ_t (or its logarithm) is often represented by another stochastic differential equation. Given a stochastic process for asset prices, we define *integrated volatility* over the period $[t, t + \Delta]$ as

$$\int_t^{t+\Delta} \sigma_s^2 \, ds.$$

It is known that realized volatility converges in probability to integrated volatility as the time split π become finer (i.e., when realized volatility is calculated using higher frequency intraday data). This convergence holds under very mild conditions on σ_t. For instance, σ_t needs be neither an Itô process nor a Markov process, and this convergence holds in case of fractional Brownian motion as well. As such, realized volatility is not only easily calculated from data without relying on a specific model, but also has the theoretically desirable property that it converges to integrated volatility under mild conditions as described above. These advantageous points have contributed to the widespread use of realized volatility as a benchmark HV measure in financial econometrics, especially since the 2000s.

Realized volatility is also known to have long memory, and many analyses have used the ARFIMA model to describe realized volatility.[24] Recently, however, the HAR (Heterogeneous Autoregressive) model proposed by Corsi [2008] and its extensions have been widely used to describe realized volatility. The HAR model predicts future realized volatility based on a mixture of realized volatility measured over multiple time horizons. For example, a future daily realized volatility is predicted based on its lag and historical realized volatility measured at weekly and monthly frequencies. Although the HAR model has a simple model structure and does not have long-memory property, it is widely used because of its high predictive ability for daily realized volatility. Similar to the GARCH model, there exist extensions of the HAR model such as an asymmetric HAR model that takes into account

[23] In, for example, Andersen et al. [2001], the asset price process is a general semimartingale. We omit the jump term for simplicity in this paper. For the properties of realized volatility and consistent estimators in the presence of a jump term, see Barndorff-Nielsen and Shephard [2004].

[24] There is an alternative approach based on a state-space model called the unobserved components model. See, for example, Nagakura and Watanabe [2015].

leverage effects (e.g., Ubukata and Watanabe [2014], Bekaert and Hoerova [2014]), and HAR-CJ models that consider jumps in the asset pricing process (Andersen et al. [2007]).

Finally, we discuss the empirical properties of the correlation between realized volatility and asset returns, called the *Zumbach effect* (Blanc et al. [2017], El Euch et al. [2020]). First, we denote the cross-correlation function between the daily integrated variance and the squared asset returns as follows:

$$\widetilde{C}^2(\tau) = E\left[\left(\sigma_t^2 - E\left[\sigma_t^2\right]\right) r_{t-\tau}^2\right].$$

Using this, we introduce the following quantity called time-reversal asymmetry (TRA):

$$Z(\tau) = \widetilde{C}^2(\tau) - \widetilde{C}^2(-\tau), \quad \tau > 0.$$

In words, TRA measures the difference between the covariance when volatility precedes the squared return and the covariance when the squared return precedes volatility. A series of studies by Zumbach and his co-authors have revealed that when TRA is calculated using the realized variance as an estimate of the integrated variance, TRA is positive for many financial time series (e.g., Zumbach and Lynch [2001], Lynch and Zumbach [2003], Zumbach [2004], Zumbach [2009]). On the other hand, one can show theoretically that TRA is (nearly) zero under standard continuous-time volatility models. This implies that the Zumbach effect cannot be represented by standard models. Recent studies on rough volatility have shown that the Zumbach effect can be represented by a rough volatility model. This point is discussed in section 4.

3.2 Overview of Gatheral et al. [2018]

As mentioned above, there has been a long history of theoretical and empirical studies on the volatility of asset prices. Volatility models have been developed so that they capture the time-series features of observed data such as long memory as the most prominent example. A noteworthy trend in recent years is the development of an analytical framework based on realized volatility calculated from intraday data. Concurrently, the 2000s saw progress in research that suggests the need for roughness based on observations about IVs (see section 4 for details). Motivated by studies on IVs, Gatheral et al. [2018], the working paper version of which was published in 2014, analyzed the roughness of realized volatility. In that study, they revealed that the realized volatility of equity indices and several other asset prices has a roughness property, which attracted a great deal of attention. In the following, we provide an overview of Gatheral et al. [2018].

Using data on realized volatility calculated from intraday data for various financial assets, Gatheral et al. [2018] estimated the degree of roughness H of their paths (i.e., time series of realized volatility). This roughness measure corresponds to the Hurst index of fractional Brownian motion, and thus $H < 1/2$ suggests that the variation of realized volatility is rougher than that of a Brownian motion. In the remainder of this subsection, we discuss roughness with the estimation of the Hurst parameter of fractional Brownian motion in mind.[25]

As we have seen in section 2.2, when volatility follows a fractional Brownian motion with Hurst parameter $H \neq 1/2$, if the interval Δ over which the volatility is measured is varied, its standard deviation does not satisfy the so-called *square-root-of-time rule*. More specifically, when the logarithm of realized volatility follows a fractional Brownian motion with Hurst parameter H, the following equation holds with respect to the absolute value of the variance of realized volatility from (4)(iii) in section 2.2, where the left-hand side of the lower equation is the (generalized) moments with q being a non-negative real number representing a power exponent:

$$E\left[\left|\log \sigma_{t+\Delta} - \log \sigma_t\right|^q\right] = K_q \Delta^{qH}. \tag{7}$$

Taking the logarithm of both sides, $\log\left(E\left[\left|\log \sigma_{t+\Delta} - \log \sigma_t\right|^q\right]\right) = const + qH \log \Delta$. Provided that this relationship holds, if we fix q and calculate the expectation on the left-hand side of equation (7) for various intervals Δ, then we can expect a log-log plot of this left-hand side versus Δ to be a straight line with slope qH.

Gatheral et al. [2018] empirically examined this conjecture based on the daily realized volatility calculated from five-minute data for DAX futures, German government bond futures, the S&P 500 index, and the NASDAQ index. They first calculated the logarithm of the left-hand side of equation (7) by replacing the expectation with the sample mean with five different q ($q = 0.5, 1, 1.5, 2, 3$) and Δ = one day to one year. Then, for each fixed q, they regressed the left-hand side on $\log \Delta$ to obtain an estimate of qH. Next, they obtained an estimate of the Hurst parameter H by regressing the estimates of qH obtained above on q. They reported that the estimated values of the Hurst parameter to be stable at around 0.1 (0.082 − 0.142) regardless of the asset type among the above four, with the fit of their regressions being extremely good. The fact that the Hurst parameter was estimated stably below 1/2 suggests that the volatility of these assets exhibits rough stochastic dynamics.

To explain these empirical findings consistently, Gatheral et al. [2018] proposed the Rough Fractional Stochastic Volatility (RFSV) model. In the RFSV model, the instantaneous volatility is given as $\sigma_t = \exp(X_t)$, where X_t is a fractional Ornstein-Uhlenbeck (fOU) process,

$$dX_t = \nu dB_t^H - \alpha (X_t - m)\, dt,$$

in which B_t^H is a fractional Brownian motion with Hurst parameter H, and ν, α, m are constant parameters. The mean reversion parameter α is assumed to be sufficiently small that the contribution of the mean reversion term (the dt term in the above equation) is negligible on short time scales. Under this assumption on α, it has been shown that X_t behaves locally like a fractional Brownian motion, yet it is a stationary process (see Proposition 3.1 in Gatheral

[25] As mentioned in section 2.2, roughness is defined based on the Hölder continuity of paths. Therefore, the degree of roughness can be defined and measured even when volatility follows a stochastic process other than a fractional Brownian motion. Nevertheless, a majority of rough volatility research conducts analyses with fractional Brownian motion in mind. For the sake of the simplicity of exposition, our discussion also has the estimation of the Hurst index of fractional Brownian motion in mind.

et al. [2018] for the precise mathematical description). For fractional Brownian motion and RFSV models to be consistent with the empirical findings in Gatheral et al. [2018], the Hurst parameter for fractional Brownian motion in RFSV models must be less than $1/2$. Conversely, Gatheral et al. [2018] reported that a Comte and Renault [1998]-type long memory model with $H > 1/2$ does not fit to the observations on the behavior of realized volatility variations.

Gatheral et al. [2018] also compared the accuracy of volatility forecasts based on the RFSV model with existing models such as AR and HAR models, and reported that the RFSV model showed better forecasting accuracy than AR and HAR models for almost all stock indices and volatility measurement horizons. They also showed that only one parameter, the Hurst index H, needs to be estimated to forecast volatility.[26] They argued that this feature of the RFSV model is a practically appealing point, as it requires estimating much fewer parameters than AR and HAR models.[27]

These results of Gatheral et al. [2018] have attracted much attention from researchers and practitioners. One reason for this is the relationship between the long memory property mentioned in section 3.1.3 and roughness. For example, fractional stochastic volatility (FSV) models (e.g., Comte and Renault [1998]) have been used as continuous-time volatility models with long memory. Since FSV models require the Hurst index to be greater than $1/2$, strong academic interest has emerged on the relationship between these models and the results reported by Gatheral et al. [2018]. In addition, practitioners also paid great attention to rough volatility due to its superior volatility forecasting ability. Progress in research on the empirical and theoretical properties of IV and its relationship to rough volatility, which will be discussed in detail in section 4, has also fueled interest in rough volatility.

3.3 Recent research trends in roughness of HV

3.3.1 Measurement error in realized volatility and estimation of the Hurst parameter

Recent research following Gatheral et al. [2018] has mainly dealt with two issues. The first issue is the estimation methods for the Hurst parameter. Gatheral et al. [2018] used daily realized volatility calculated from five-minute data as input data and estimated the Hurst parameter based on a regression analysis based on equation (7). In this estimation, the

[26]The RFSV model has the mean reversion parameter α, as well as other parameters, in addition to the Hurst parameter. However, as noted above, α does not appear in the volatility forecasting formula, because of the assumption that the mean reversion term can be ignored on short time scales. Gatheral et al. [2018] showed that the other parameters also do not appear in the forecasting formula, because they cancel each other out in the derivation process.

[27]Note that Gatheral et al. [2018] simply compared the accuracy of predictions of each model without conducting statistical tests regarding the differences between them. Wang et al. [2023a] compared forecasting accuracy between rough models and non-rough models (e.g., HAR models) based on the Diebold and Mariano [2002] test and the model confidence set of Hansen et al. [2011], and found that the fOU process, which is a rough model, has significantly higher predictive ability. In addition, Wang et al. [2023b] analyzed an optimal way in terms of prediction accuracy to approximate the expected value of fractional Brownian motion with their discretely sampled values.

measurement errors of daily realized volatility are not taken into account; it is (implicitly) assumed that the input data coincide with the true volatility measure.

Note, however, that realized volatility differs from true instantaneous volatility σ_t and its cumulative value over a given period, integrated volatility $\int \sigma_s \, ds$, because realized volatility is calculated from discretely sampled asset price data.[28] Thus, a naive estimation approach could lead to bias in the estimated value of the Hurst parameter, due to possible measurement errors in the input data. In fact, Fukasawa et al. [2022] estimated the Hurst parameter based on a regression analysis using daily realized volatility calculated from five-minute data as input, and found that the Hurst parameter (as Gatheral et al. [2018] reported) takes a value around 0.1 regardless of the true Hurst parameter for the data-generating processes. They argued that measurement errors in realized volatility drive this result. Cont and Das [2023] also reported that estimating the Hurst parameter based on the variation of the path of realized volatility leads to biased estimates. Based on this, they pointed out that obtaining an estimated Hurst parameter less than 1/2 does not necessarily serve as evidence of roughness.

In response to this issue, several papers have proposed parametric estimation methods for the Hurst parameter by introducing a model of measurement errors in realized volatility. The aforementioned Fukasawa et al. [2022] proved a limit theorem on the difference between realized and integrated volatility (i.e., measurement error of realized volatility) under a mild assumption that the logarithmic process of the underlying asset follows a continuous semimartingale. Specifically, they showed that the measurement error of realized volatility follows a normal distribution with mean 0 and variance $2/m$ (where m is the number of intraday data per day). They then proposed a quasi-maximum likelihood estimator based on this theoretical result. When this estimator is applied to the five-minute realized volatility of major equity indices such as the S&P 500 and the Nikkei 225 Index, the estimated Hurst parameters are around 0.02-0.06, which are smaller than the Hurst parameters reported in Gatheral et al. [2018]. Similar to Fukasawa et al. [2022], Bolko et al. [2023] proposed a generalized method of moment (GMM) approach to estimate the Hurst parameter under the assumption that the logarithmic process of the underlying asset follows a continuous semimartingale. Their empirical analysis also suggested that the time evolutions of the volatilities of major equity indices are very rough, with estimated Hurst parameter values below 0.05.

Damian and Frey [2023] proposed a model wherein high-frequency asset price fluctuations are described by a point process and the intensity of the point process is determined by latent factors that correspond to volatility.[29] In this setting, the authors formulated the estimation of volatility as a filtering problem of point process data. However, since fractional Brownian motion is not Markovian and thus the behavior of stochastic processes is path-dependent, using conventional filtering techniques is not realistic due to computational

[28] Jumps in the underlying price process also create gaps between realized volatility and integrated volatility, because realized volatility is affected by jumps, whereas integrated volatility is not. See Andersen et al. [2012] and Barndorff-Nielsen [2005] for the estimation of realized volatility when the underlying process may jump.

[29] A point process is a stochastic process that describes randomly distributed points (events) over time or space, such as the number of contingent events, and intensity controls the likelihood of the occurrence of events. See section 5.2 for more details on point processes.

burden, including the large memory requirements. To overcome this difficulty, Damian and Frey [2023] transformed fractional Brownian motion to superposition of an infinite number of OU processes. Because an OU process is an Itô process, this transformation renders the application of a particle filtering method possible, because it requires the Markov property. They performed various simulation analyses and reported that their approach is able to distinguish true roughness from the spurious roughness caused by the presence of noise, as pointed out by Cont and Das [2023].

A topic related to measurement errors is *market microstructure noise* (MM noise) in observed asset prices. Typical examples of MM noise include rounding errors caused by discrete tick sizes at exchanges (e.g., one cent minimum tick size) and the bid-ask bounce, that is, mechanical negative autocorrelations due to the existence of bid-ask spreads.[30] When an underlying asset price contains MM noise, estimation of volatility using high-frequency data requires caution. As seen in section 3.1, the property that the quadratic variation of the underlying asset price converges to the integrated volatility as the time split becomes finer, which is the theoretical basis for using realized volatility, holds under the assumption of no MM noise in the underlying asset price. However, if the underlying asset price contains MM noise, it can be shown that realized volatility converges to the variance of the MM noise, not to the integrated volatility. To avoid this problematic result, appropriate treatment is required (Zhou [1996], Aït-Sahalia and Mykland [2009]).

Several approaches exist for dealing with MM noise in high-frequency data. The simplest and most widely used approach recommended in the literature (Andersen et al. [2001], Gençay et al. [2002], Barndorff-Nielsen and Shephard [2002], among others) is down-sampling, that is, sampling data infrequently (e.g., five-minute data) so that the effects of MM noise in high-frequency data can be mitigated.[31] In addition to mitigating MM noise, down-sampling has the advantage that it is easy to handle both theoretically and empirically because it is estimated from equidistant time-series data (unlike unevenly distanced tick data). Many studies have shown that calculating realized volatility from five-minute data is very simple and convenient, yet this method strikes a good balance between having a sufficiently high data frequency and obtaining sufficient mitigation of MM noise. On these grounds, Fukasawa et al. [2022] also conducted an analysis using realized volatility calculated from five-minute data (Fukasawa et al. [2022], Remark 2.4 (ii)).

[30]For an overview of the impact of market microstructure on asset prices, see Campbell et al. [1997].

[31]Bandi and Russell [2008] provided an estimate of the optimal down-sampling data frequency. Note that down-sampling is in one sense suboptimal, as it discards a portion of high-frequency data, which may contain useful information. For this reason, alternative approaches have been proposed to utilize high-frequency data as a whole by taking the existence of noise into account. For example, Zhang et al. [2005] proposed a method called Two Scales Realized Volatility, which simultaneously estimates the integrated volatility and variance of MM noise in a consistent manner. There are also estimations based on state-space models (e.g., Nagakura and Watanabe [2015]). Hansen and Lunde [2006] analyzed the effect of MM noise in the measurement of realized volatility and proposed a realized volatility indicator that corrects for the effect of MM noise. Nevertheless, some studies have suggested that down-sampling is the best approach for practical applications, given the fact that it is a simple method yet it achieves sufficiently good performance. In fact, Liu et al. [2015] compared more than 400 estimates of realized volatility and reported that no estimation methods clearly outperform five-minute realized volatility.

Note that the discussion on MM noise should not be confused with that on measurement errors in Fukasawa et al. [2022] and others, which analyzed the impact of measurement errors in realized volatility calculated from discretely sampled asset prices on the estimation of the Hurst parameter, as opposed to integrated volatility defined from a continuous-time stochastic process of volatility. Such measurement errors can occur even in the absence of MM noise. Therefore, bias in the estimates of the Hurst parameter from regressions of realized volatility, as done by Gatheral et al. [2018], can occur even in the absence of MM noise.

3.3.2 Relation to the long memory property

The second issue regarding the roughness of realized volatility is the relationship between the long memory property and roughness. Fractional Brownian motion (and standard Brownian motion) has a property called self-similarity (see section 2) and thus rescaling the time scale does not change its distributional properties once an appropriate scaling coefficient is attached. In other words, the behavior of fractional Brownian motion over very short time horizons and that over very long time horizons are essentially identical. This means that long memory (positive serial correlation over a long period of time) and roughness (negative serial correlation due to wild fluctuations over a short period of time) cannot be represented by a single fractional Brownian motion. There are two directions to trying to address this issue. One strand of studies has raised the possibility of *spurious* detection of long memory, while the other strand has aimed to develop a model that can represent roughness and long memory simultaneously.

The existing literature on long memory has shown that, even when the true data-generating process does not have long memory, estimated long memory parameter (fractional integration parameter) d may be biased in the presence of time trends, structural breaks, regime changes, and the like, leading to spurious detection of long memory (e.g., Diebold and Inoue [2001]; Granger and Hyung [2004]). Gatheral et al. [2018] also discussed the possibility of spurious detection of long memory; they reported that when the Andersen et al. [2001] method for estimating the fractional integration parameter d of the ARFIMA model is applied to the time-series data generated by the RFSV model, the estimation result yields spurious detection of long memory.[32] Based on this observation, they argued that the RFSV model, like those of the previous studies mentioned above, is a time-series model that is rough in nature but can be estimated spuriously to have long memory.

Another strand of literature proposes models that can represent roughness and long memory simultaneously. For example, Bennedsen et al. [2021] proposed a model using a stochastic process called a *Brownian semi-stationary process* (Barndorff-Nielsen and Schmiegel [2007], Barndorff-Nielsen and Schmiegel [2009]), which is an extension of Brownian mo-

[32] The Andersen et al. [2001] estimation method examines the Fourier transformation of the autocorrelation function to see whether it satisfies a condition corresponding to a hyperbolic decay of the autocorrelation function in the time domain.

tion. In their model, the roughness parameter and the autocorrelation coefficients of long-distance lags are distinct, allowing the degree of roughness and long memory to set independently. Furthermore, they also proposed an empirical method to estimate this model based on a nonlinear least squares method while taking into account measurement errors in realized volatility. Applying their empirical method to the S&P 500 index futures and individual stocks data, they obtained results which suggest that the volatility of these assets has both the rough property and the long memory property.

3.3.3 Summary of this section and future prospects

To summarize the current research trends on the roughness of realized volatility, the following points remain key issues. First, realized volatility contains measurement errors and differs from its theoretical counterpart, integrated volatility. Therefore, when realized volatility is used as input data as in Gatheral et al. [2018], estimated values of the Hurst parameter may suffer from a downward bias (i.e., bias to the direction of roughness) unless this measurement issue is appropriately taken into account. For this purpose, several recent studies proposed more sophisticated estimation methods that take this issue in account. These studies reported that the volatility of equity indices and individual stocks exhibit even stronger roughness according to these improved estimation methods.

Second, due to the self-similarity of fractional Brownian motion, at least simple models such as a single fractional Brownian motion cannot possess roughness and long memory simultaneously. There exist two main arguments on this issue. First, some studies have raised the possibility that the long memory property may be spuriously detected. For example, Gatheral et al. [2018] demonstrated that a conventional approach spuriously detects long memory even when it is applied to time series generated by the RFSV model. Another strand of studies proposed stochastic processes that can simultaneously express roughness and long memory.

Both the above issues exemplify the difficulty of analyzing the nature of (instantaneous and integrated) volatility, which is not directly observable. This calls for continued development of less noisy estimates of volatility, more realistic models that can describe the complex features of volatility, and more robust statistical methods for detecting roughness and long memory.

Finally, we note that some researchers have expressed skepticism about the need for complex models based on fractional Brownian motion. For example, Rogers [2023] questioned the rough volatility model based on fractional Brownian motion for several reasons. First, he pointed out that its non-Markovian nature prevents it from practical applications due to its heavy computational cost, which includes its memory usage cost. Second, he doubted that there exists a convincing financial economic mechanism that yields non-Markovian asset price dynamics. Third, he argued that a conventional Itô process seems to be sufficient to describe the time-series characteristics over typical time scales in empirical research (e.g., daily or weekly frequency), suggesting limited empirical benefit from using complicated fractional Brownian motion-based models. Thinking deeper about these issues may require holistic judgment, considering not only the perspectives arising from HV research (i.e., observations about realized volatility and estimation methods for the degree of roughness) but also other perspectives. Specifically, our subsequent sections provide additional perspectives, such as the necessity of roughness in derivative pricing and for risk management purposes, as well as the financial economic micro-foundations for explaining why roughness arises.

4 Shape of implied volatility surface and rough volatility models

In this section, we first provide an overview of the development of mathematical finance from both practical and academic perspectives, and then discuss the roughness of volatility indicated by the shape of IV (particularly, its term structure), why we need rough volatility models for pricing derivatives, and prospects for their practical applications. Mathematical finance and financial engineering[33] have mainly focused on pricing and risk management of derivatives and have provided a theory to calculate fair prices using Itô calculus (stochastic differential equations) under the principles of no-arbitrage and replication.[34] Interrelationship between theory and practice has progressed these fields; application of theory has helped solving practical problems and problems arising from practice issues have stimulated the further development of the theory.

The *IV representation problem*, on which we focus in this section, is a typical example of such progress. Rough volatility initially attracted attention as a method that could solve this problem, especially the expression of the *negative power law* regarding the term structure of IV, which was considered difficult to solve with existing methods. This discovery led to subsequent research from the perspective of HV and MM. In recent years, *rough volatility models* for derivative pricing have also been developed. In particular, it was shown that, despite its non-Markovian nature, a rough volatility model can be regarded as a Markov model by considering additional variables such as forward variances observed in the market. This result is referred to as the recovery of the Markov property and allows us to use the existing framework of mathematical finance theory for rough volatility models. Furthermore, a recent theoretical result showed that under markets where roughness is implied, arbitrage opportunities will arise if derivatives are not priced and their risk is not managed by rough models. This result is of importance as it necessitates the use of rough models. In addition, it has been found that a rough volatility model can solve the *joint calibration problem of SPX*

[33] Both mathematical finance and financial engineering focus on the pricing of derivatives. The former research field is often referred to as a branch of applied mathematics mainly focusing on the fair valuation of derivatives based on martingale theory, whereas the latter is often referred to as a research field oriented towards practical applications, mainly focusing on practical uses of models such as matching model parameters to reproduce market observations.

[34] No-arbitrage is the condition that one cannot make a profit for sure without incurring an initial cost. Replication refers to generating the same cash flow as the derivative by appropriately trading the underlying asset and other assets.

options' IV and VIX options' IV, which has been considered difficult to solve with existing derivatives models.

Section 4.1 summarizes the development of mathematical finance, by focusing on the IV representation problem while taking into account practical issues. Section 4.2 introduces rough volatility studies related to IV. Specifically, we explain that a certain shape of IV indicates the roughness of volatility and the need for rough volatility models. Then, we summarize how rough volatility models can be used for derivative pricing and risk management, as well as providing prospects for the further development of rough volatility models.

4.1 History of the development of mathematical finance and financial engineering: IV representation problem

4.1.1 Implied volatility

As mentioned previously, the Black–Scholes model (Black and Scholes [1973]) is the most basic model and has been the key driver of the rise of mathematical finance and financial engineering, as well as the development of derivatives markets. As will be discussed shortly, the Black–Scholes model serves as a basic tool in current practice of financial institutions, for example, when calculating IVs given the market prices of options.

In the Black–Scholes model, the underlying asset price is assumed to follow a log-normal distribution, which allows for analytical pricing of vanilla options such as European call options (i.e., the pricing formula is known and the price can be calculated without numerical approximation). Here, the option price can be calculated as the expected value of the payoff under a risk-neutral probability measure, a probability measure where the expected return on the underlying asset price is equal to the risk-free interest rate. As a result, the calculated price is equal to the initial cost required to replicate this option under the no-arbitrage assumption (for details, see, for example, Björk [2009]). Therefore, by using the Black–Scholes pricing formula, the option price can be obtained without estimating the expected underlying return under the physical probability measure.

Formally, the Black–Scholes pricing formula is formulated as follows. Let W_t^Q be a Brownian motion under a risk-neutral probability measure and let the underlying asset price $S_t^{(\mathrm{BS})}$ follow a geometric Brownian motion (i.e., the price at a certain time follows a log-normal distribution):

$$\frac{dS_t^{(\mathrm{BS})}}{S_t^{(\mathrm{BS})}} = r\,dt + \sigma\,dW_t^Q, \quad \text{i.e., } S_t^{(\mathrm{BS})} = S_0^{(\mathrm{BS})} \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t^Q\right),$$

where t is the current time, r is the risk-free rate, and σ is the volatility parameter. Moreover, let K be the strike and T be the maturity of a European call option. Then, the price of this option is given as follows:

$$
\begin{aligned}
C_{\mathrm{BS}}(r, K, T, \sigma, S_t) &= E^Q\left[e^{-r(T-t)}\max\{S_T^{(\mathrm{BS})} - K, 0\} \mid \mathcal{F}_t\right] \\
&= S_t^{(\mathrm{BS})}\Phi\left(\frac{\log\left(\frac{S_t^{(\mathrm{BS})}}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right) - e^{-r(T-t)}K\Phi\left(\frac{\log\left(\frac{S_t^{(\mathrm{BS})}}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right),
\end{aligned}
$$

where Φ is the standard normal cumulative distribution function.

The parameters (arguments) that appear in the Black–Scholes pricing formula are the current underlying asset price S_t, the risk-free rate r, the volatility parameter σ, the strike K, and the maturity T. Since the underlying asset price and the risk-free rate are observable in the market[35] and the strike and maturity are terms of the instrument being priced, σ is the only free parameter. This means that there is one-to-one relationship between option prices and the volatility parameter σ. As a result, the corresponding volatility can be backed out from a given option price using the Black–Scholes formula.[36] Specifically, given a (market) price $C^{(\mathrm{mkt})}(K, T)$ of a European call option whose strike is K and maturity is T, the Black–Scholes implied volatility (BS-IV) is implicitly defined based on the following equation:[37]

$$
C^{(\mathrm{mkt})}(K, T) \quad = C_{\mathrm{BS}}\left(K, T, \sigma_{\mathrm{BS}}^{(\mathrm{mkt})}(K, T)\right).
$$

In practice, option prices are conventionally quoted in terms of IVs.[38] This convention is based on the idea that the market prices of options reflect market expectations on future uncertainty.

4.1.2 Volatility Surface

If the underlying asset price follows the Black–Scholes model, BS-IVs take a constant value equals to the constant volatility parameter σ regardless of maturity or strike. On the other hand, BS-IVs calculated from actual market prices of options usually take different

[35] The risk-free rate is typically calculated from bond or swap prices.

[36] Since the function C_{BS} in the Black–Scholes formula is a monotonically increasing function with respect to σ, the implied volatility obtained from the inverse function is uniquely determined.

[37] There are other types of IVs. For example, Black (Bachelier) IV is calculated assuming that the underlying price process follows a Brownian motion, and is used in the options market where the underlying asset price may become negative (e.g., interest rates). Another prominent type of IV is the so-called *model-free IV*, which is calculated without assuming a specific model for the underlying asset price dynamics. As described below, the calculation methodology of volatility indices such as VIX is grounded on the theory of model-free IV.

[38] Since option prices fluctuate according to the price of the underlying asset, the fluctuation of the option price is volatile if its underlying asset price is volatile (e.g., exchange rates). Therefore, it would be cumbersome from a practical standpoint if options were indicated in terms of price, because option prices will change whenever the price of the underlying asset changes. In addition, since option prices also depend on maturity and strike, option prices with different maturities and strikes are not directly comparable. Therefore, the market practice of displaying the value of options in terms of (standardized) volatilities that are independent of the underlying price, maturity, and strike has been widespread. Note that not all derivative instruments display their value in terms of volatility; for example, foreign exchange options on minor currencies are sometimes quoted in terms of prices.

values depending on strike K and maturity T.[39] Namely, given the market prices of options for various strikes and maturities, the BS-IVs depicted as a three-dimensional surface as a function of strike K and maturity T is called a *volatility surface*. In practice, it is conventional to use log-moneyness $\kappa = \log\left(\frac{e^{-r(T-t)}K}{S_t}\right)$ instead of strike K, and time-to-maturity $\tau = T - t$ instead of maturity T as arguments. Figure 3 shows a typical volatility surface for, say, an equity options market.

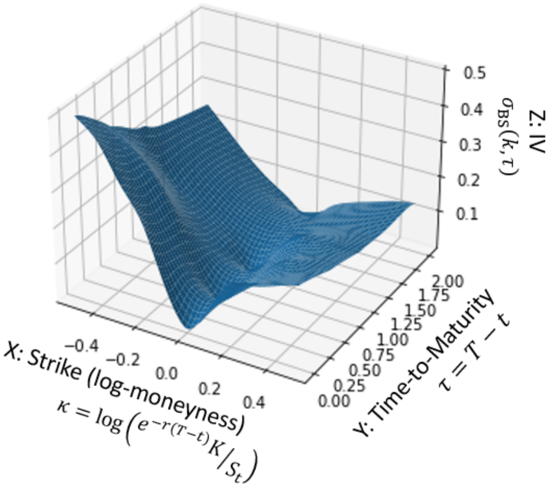

Figure 3: Volatility surface (conceptual drawing)

In the equity index options markets, which are among the most liquidly traded options markets, the market prices of options with highly liquid strikes and maturities are often quoted as volatility surfaces. Looking at the cross-section of the volatility surface at a single maturity τ, we see that IVs tend to be higher for lower strike prices. This corresponds to a thicker distribution in the direction of falling stock prices, a phenomenon known as *volatility smile/skew*. Looking at the cross-section at a single strike price (cross-section along the Y-axis), we see that IVs tend to be higher for shorter maturities. This corresponds to the fact that the prices of options with shorter maturities are more heavily affected by the risk of sudden changes in the market.

4.1.3 Representation of IVs

IV representation refers to structuring derivative pricing models[40] so that they are consistent with the volatility surfaces quoted in the market. The motivation and details are

[39] This means that although the Black–Scholes model is used in calculating BS-IV, the underlying asset price does not follow the Black–Scholes model under a risk-neutral probability measure.

[40] Derivative pricing models can be categorized into vanilla models for pricing relatively simple instruments (vanilla options) and exotic models for pricing complex instruments (exotic derivatives). In both cases, the market prices of highly liquid instruments are used as input to calculate the price of instruments to be priced. For vanilla options, the main calculation performed in the vanilla model is only the interpolation of the market prices as input, since similar instruments (e.g., the same derivative products with different maturities) have high market liquidity. On the other hand, as discussed in the text, the exotic model requires more complex

described below. First, note that the general flow of derivative pricing process in practice is as follows:

(1) Build a model to represent the trading prices of highly liquid derivative instruments quoted in the market.

(2) Using the model built in step (1), price and hedge more complicated derivative instruments.

Here, step (1) is necessary to build a model that accurately represents the prices of highly liquid derivatives and calculate the prices of more complex derivatives while satisfying the no-arbitrage condition.[41] Since there is one-to-one relationship between option prices and BS-IVs as mentioned above, it is important to construct a model that can represent IVs of highly liquid derivatives (e.g., European options). Accordingly, it has been a central theme of mathematical finance to construct models that can represent the features of volatility surfaces observed in the market (i.e., construct a stochastic differential equation that the underlying asset price follows), as well as efficient methods to adjust the parameters of the model to reproduce the observed market prices, which is referred to as *calibration*. Regarding step (2), the calculation of the prices of complex derivatives and derivation of hedging strategies (i.e., the derivation of the sensitivity of option prices to changes in underlying asset prices, etc., called *Greeks*) are frequently required in practice. Therefore, it is desirable to perform these calculations with high accuracy and speed. In addition, when deriving hedging strategies, it is important to capture not only the volatility surface at a single time point, but also its dynamics (time evolution). There have been various studies on these issues as they are important themes in theoretical research and practical applications of mathematical finance.

4.1.4 Existing derivative pricing models

Let us list some representative derivative pricing models: local volatility models that describe volatility as a deterministic function (Dupire [1994]), stochastic volatility models in which volatility varies stochastically (Heston [1993], Hagan et al. [2002]), and jump models in which volatility may vary discontinuously (Bates [1996]). Table 1 summarizes these models.

calculations, such as obtaining parameters under risk-neutral probability measures of the underlying asset price process. The concept of IV representation is common to both vanilla and exotic models, but our subsequent discussions primarily have the exotic models in mind.

[41] Reasons why it is desirable to maintain the no-arbitrage condition include, among others, (1) to avoid being exploited by arbitrage, (2) the need for reliable mark-to-market valuation for accounting purposes, and (3) the need to trade other liquid options for risk management purposes.

Model: BS model (Black and Scholes [1973])
Volatility term: $V_t = \sigma^2$ (Constant)
Merit: Admits analytical pricing formulae for some options (e.g., European call option)
Drawback: Observed volatilities are not constant
Model: Local Volatility model (Dupire [1994])
Volatility term: $V_t = \sigma(t, S_t)$ (Deterministic function)
Merit: Can capture volatility surface accurately at a single fixed time
Drawback: Cannot capture the dynamics of volatility surface
Model: Stochastic Volatility model (Heston [1993])
Volatility term: $dV_t = \mu^V(t, V_t)\,dt + \sigma^V(t, V_t)\,dW_t^V$ (Stochastic process)
Merit: Can capture the dynamics of volatility surface
Drawback: Cannot fully capture the shape of volatility surface
Model: Jump model (Bates [1996])
Volatility term: $dV_t = \mu^V(t, V_t)\,dt + \sigma^V(t, V_t)\,dW_t^V + dJ_t$ (Stochastic process with jump dJ_t)
Merit: Can capture the dynamics of volatility surface flexibly thanks to the jump term
Drawback: Hedging involves costly and unstable calculations

Table 1: Summary of representative derivative pricing models

4.1.5 Calibration and asymptotic expansion of IV

Let us assume that $\left\{S_u^{(\text{model})}\right\}_{t \le u \le T}$ follows the following stochastic differential equation under the risk-neutral probability measure:

$$\frac{dS_t^{(\text{model})}}{S_t^{(\text{model})}} = r\,dt + \sqrt{V_t}\,dW_t^Q.$$

Then, calibration means the calculation of the model parameters so that the following approximation holds:

$$C_{\text{BS}}\left(r, K, T, \sigma, S_t^{(\text{model})}\right) \approx E\left[\max\left\{S_T^{(\text{model})} - K, 0\right\}\right], \tag{8}$$

where $\hat{\sigma}_{\text{BS}}^{(\text{mkt})}(K, T)$ is (interpolated) IV observed in the market.

When the expectation on the right-hand side of equation (8) cannot be calculated analytically, it is common to use an approximation method such as asymptotic expansion to address this calibration problem. Highly accurate asymptotic expansion methods are necessary to obtain sufficient accuracy, and have been an important research topic in mathematical finance (see, for example, Hagan et al. [2002] and Osajima [2007]).

4.2 Progress in using rough volatility

Below, we first summarize the history of rough volatility studies related to IV, and then describe the main technical points of these studies in section 4.2.1 and beyond.[42]

As already discussed, derivative pricing models have developed toward using more complex volatility functions, motivated by the representation problem of volatility surfaces and risk management associated with their time evolution. In the early 2000s, empirical studies on the equity derivatives markets reported a characteristic pattern called *the negative power law* of volatility surfaces, which is difficult to express by existing derivative pricing models. In response to this, in the late 2000s, Alòs et al. [2007] and Fukasawa [2011] showed that the negative power law can be represented by using a model based on fractional Brownian motion. This result, together with the subsequent observation of roughness in HVs, triggered a surge of interest in rough volatility research.

In subsequent studies, rough volatility models for pricing derivatives were proposed, such as the rough Bergomi model (Bayer et al. [2016]) and the rough Heston model (El Euch and Rosenbaum [2019]). These models had been considered to have the difficulty of being non-Markovian in nature, due to fractional Brownian motion, which makes the existing framework of mathematical finance inapplicable. In regard to this, subsequent research has shown that the Markov property can be recovered by including information on the forward variance (expected value of the variance at a future time). It has been shown that, if such information is included, the theory of mathematical finance based on replication and no-arbitrage can be applied, and thus hedging strategies can be constructed in a practically feasible manner (El Euch and Rosenbaum [2018], Fukasawa et al. [2021]). One of the practical advantages of a rough volatility model is that it can represent the negative power law of the volatility surface observed in the market, despite the relatively small number of model parameters. In addition, Gatheral et al. [2020] showed that the rough volatility model can solve the *joint calibration problem of SPX options' IVs and VIX options' IVs* (the problem of simultaneously fitting to both the IVs of the S&P 500 index options and that of the VIX options in a unified model), which was considered difficult to solve with existing derivatives models.

In addition to these developments, Fukasawa [2021] proved that, in a general setting, arbitrage opportunities arise if we do not use a pricing model with appropriate roughness, as indicated by the shape of the volatility surface observed in the market. In other words, in markets where the negative power law is observed, pricing and risk management of derivatives using a conventional non-rough volatility model is undesirable because arbitrage opportunities will arise; that is, *volatility has to be rough*. This is a very important theoretical result that strongly suggests the need to use rough volatility models in the pricing of derivatives.

4.2.1 Negative power law and its representation via fractional Brownian motion

Carr and Wu [2003], Lee [2005], and Fouque et al. [2003] observed the negative power

[42] See Alòs and León [2021] and Di Nunno et al. [2023] for surveys.

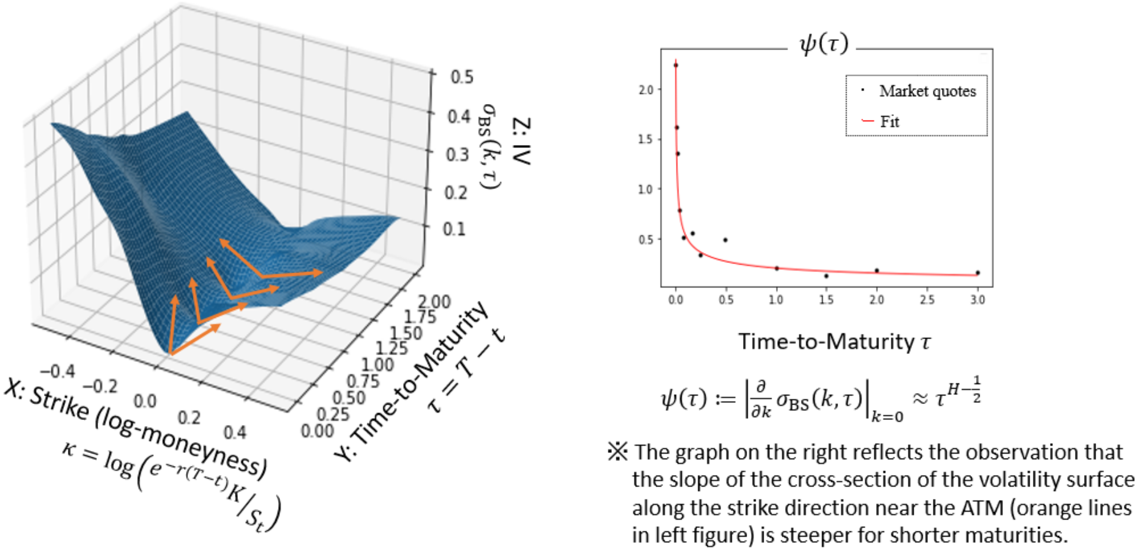

Figure 4: Negative power law (conceptual drawing)

law as a stylized fact about the shape of volatility surfaces based on real data for market prices of options. The negative power law refers to a specific shape of the term structure of IV, which is formulated as follows:

$$\psi^{(\mathrm{mkt})}(\tau) = \left|\frac{\partial}{\partial \kappa}\sigma^{(\mathrm{mkt})}_{\mathrm{BS}}(\kappa,\tau)\right|_{\kappa=0} \approx c\tau^{H-\frac{1}{2}} \quad \text{for sufficiently small } \tau > 0,$$

where $\psi^{(\mathrm{mkt})}(\tau)$ is the slope of the at-the-money (ATM) volatility skew for options with a fixed maturity τ. Namely, the negative power law means that, as a function of maturity, the slope of the volatility surface at each maturity can be described by a two-parameter function of time-to-maturity to the power of $H - 1/2$ and a constant coefficient c. It is empirically observed that $H < 1/2$ holds, namely that the slopes of the volatility surface are steeper for shorter maturities. The right graph of Figure 4 shows the ATM slopes of a typical volatility surface at various maturities (black dots) and the curve that fits these ATM slope data (red line, described by two parameters), as observed in a typical example, an equity derivatives market.

The existing Brownian motion-based models introduced in section 4.1.4 are known to have difficulty in representing the negative power law.[43] On the other hand, as mentioned above, Alòs et al. [2007] and Fukasawa [2011], as well as Fukasawa [2017], proved that the negative power law can be represented using models based on fractional Brownian motion with Hurst parameter $H < 1/2$.[44] This fact suggests that the IV observed in the market is

[43] For example, the stochastic volatility models have difficulty in representing the negative power law because IV is nearly constant at short maturities in these models. The local volatility models can represent the shape of IV at one point in time, but it is difficult to represent changes in the volatility surface as it evolves over time. The jump models can represent the behavior of volatility surfaces, but these models tend to be complicated and difficult to calibrate stably, and hedging operations become difficult in practice.

[44] Alòs et al. [2007] used an asymptotic expansion based on Malliavin calculus, and Fukasawa [2011] used a

rough, which made these studies pioneers in rough volatility research, appearing prior to the work of Gatheral et al. [2018].[45]

Moreover, Fukasawa [2021] showed that, unless we use a pricing model with an appropriate degree of roughness, namely that implied by the shape of the volatility surface observed in the market (the parameter H in the negative power law), arbitrage opportunities arise, at least in theory.[46] This indicates that no-arbitrage pricing requires a rough volatility model whose degree of roughness is consistent with the roughness of the volatility implied by the shape of the volatility surface. As mentioned, this result, together with that of research on HV and market microstructure (MM), establishes a theoretical foundation for adopting a rough volatility model.

Below, we provide a technical summary of Fukasawa [2021]. Fukasawa [2021] assumed that the underlying asset price process follows a continuous positive-valued martingale $S_t^{(\mathrm{model})}$, and considered the quadratic variation process of the log asset price $\left\langle \log S^{(\mathrm{model})} \right\rangle_{2,t}$. This quadratic variation corresponds to the integrated variance from time 0 to t, and its time derivative $\frac{d}{dt}\left\langle \log S^{(\mathrm{model})} \right\rangle_{2,t}$, corresponds to the instantaneous variance at time t. Additionally, Fukasawa [2021] assumes that the Hölder continuity of the stochastic process of this instantaneous variance is equal to H. It is then shown that the following relation holds as an approximation:

$$\psi^{(\mathrm{model})}(\tau) = \left| \frac{\partial}{\partial \kappa} \hat{\sigma}_{\mathrm{BS}}^{(\mathrm{model})}(\kappa, \tau) \right|_{\kappa=0} \propto \tau^{-\frac{1}{2}+H}, \quad \tau \to 0,\ H \approx 0.$$

This result implies that the shape of the volatility surface $\psi^{(\mathrm{model})}(\tau)$ exhibits a negative power law as long as the Hölder continuity of volatility satisfies $H < 1/2$. Next, let H be the exponent of the negative power law implied by the shape of the volatility surface, and let H_0 be the exponent of the Hölder continuity of the volatility of the stochastic process $S_t^{(\mathrm{model})}$ used as the pricing model. It is shown that if $H_0 > H$, that is, if the roughness of the pricing model is less than that of the observed IV shape, an arbitrage opportunity can be constructed. Note that this arbitrage opportunity is constructed using the fact that the roughness of the IV amplifies the error of the delta hedge based on the Black–Scholes model, and the arbitrage strategy is composed of a combination of very short-term call and put options.

However, the negative power law of IV described above has been observed mainly in the market of U.S. equity options with very short maturities, and the appropriateness of the data used for the observation and the interpolation/extrapolation methods is open to discussion. In particular, it has been pointed out that $\psi(\tau)$ observed from option price data has different curvatures for the long and short maturity portions, so it is necessary to use different param-

martingale expansion.

[45]Gatheral et al. [2018] is a pioneering study in terms of using fractional Brownian motion with $H < 1/2$ in mathematical finance. Note that, as mentioned in section 3, there have been previous studies using fractional Brownian motion with $H > 1/2$ to represent the long memory property (Comte and Renault [1998]).

[46]Fukasawa [2021] assumes an ideal financial market, where no market frictions such as transaction costs exist and vanilla options of any strike and maturity can be traded. Therefore, the results in Fukasawa [2021] do not necessarily imply that a viable arbitrage strategy could be constructed in a real market.

eters, depending on the range of IV maturities to be approximated. For example, from the examination using options market data for the EURO STOXX Index and DAX Index, Guyon and El Amrani [2022] found that although a two-parameter model fits well when focusing only on the very short portion (maturities of two weeks or less), when fitting to the entire IV including the longer maturity portion, three parameters were necessary to fit well to the entire IV. On the other hand, Delemotte et al. [2023] concluded that, consistent with previous studies, by testing with longer (2007-2015) U.S. stock options market data, a two-parameter function fits sufficiently well. Overall, similar to the discussion on long memory ($H > 1/2$) and roughness ($H < 1/2$) in factional Brownian motion mentioned in section 3, this strand of literature suggests that the characteristic patterns of volatility may differ between long and short time horizons. Other studies that have attempted to model the different shapes of IVs in different time horizons include Funahashi and Kijima [2017a] and Funahashi and Kijima [2017b].

4.2.2 Rough volatility model

In recent years, some research has showed that even when the volatility process follows a model with roughness, the asset price processes can be treated as Markov processes by additionally considering information on forward variances. This finding, coined as the recovery of the Markov property, was first explicitly shown in the rough Bergomi model (Bayer et al. [2016]), and since then, the rough Heston model and others have been shown to have similar properties. We give an overview of these models below.

Bayer et al. [2016] proposed a rough Bergomi model as a pricing model for derivatives. The rough Bergomi model is a rough stochastic process and is an extension to infinite dimensions of the Bergomi model (Bergomi [2004]), in which the forward variance is described by an Itô process. Specifically, it is given by the following equations:

$$\begin{cases} \frac{dS_t}{S_t} = \mu_t + \sqrt{V_t}\left(\rho \, dW_t^{(1)} + \sqrt{1-\rho^2} \, dW_t^{(2)}\right), \\ V_t = E\left[V_t \mid \mathcal{F}_u\right] \exp\left(\eta \sqrt{2H} \int_u^t (t-s)^{H-\frac{1}{2}} \, dW_s^{(1)} - \frac{\eta^2}{2}(t-u)^{2H}\right), \end{cases} \tag{9}$$

where t is the current time, u is a certain past time, V_t is the instantaneous variance, and $\left(W_t^{(1)}, W_t^{(2)}\right)$ is a two-dimensional Brownian motion. $E\left[V_t \mid \mathcal{F}_u\right]$ is the expectation of the future variance conditioned on information up to time u (i.e., $\mathcal{F}_u$), and is called the forward variance. Bayer et al. [2016] derived the rough Bergomi model, by modeling the spot variance based on fractional Brownian motion and representing forward variance by a Riemann—Liouville-type Itô integral.

Given the forward variance $\{E\left[V_{t+\theta} \mid \mathcal{F}_t\right]\}_{\theta>0}$, the rough Bergomi model can be regarded as a Markov process. Specifically, from the general derivative pricing formula $C_t = E\left[F\left(S_T\right) \mid \mathcal{F}_t\right]$, the rough Bergomi model admits the following expression:

$$C_t = E\left[F\left(S_T\right) \mid \mathcal{F}_t\right] = G\left(T-t, S_t, \{E\left[V_{t+\theta} \mid \mathcal{F}_t\right]\}_{\theta\geq 0}\right),$$

where G is a some function. By taking into account the third argument on the right-hand side of this equation (forward variance), we can adopt existing mathematical finance theory based on the Markov property and other desirable properties. Therefore, we can apply existing notions such as replication, no-arbitrage, and risk-neutral measures to the rough Bergomi model.

This is a strong advantage of the rough Bergomi model, as it opens the possibility of application to the U.S. equity volatility market, where forward variances can be estimated from the market price of volatility derivatives.[47] In subsequent studies of rough volatility models, the recovery of the Markov property by forward variances also played an important role. Consistent with the discussion in section 4.2.1, Bayer et al. [2016] showed that the SPX volatility surface, where the negative power law is observed, can be represented well by a rough Bergomi model having only a few parameters.[48] Extending this argument, asymptotic expansion formulae for IVs yielded from the rough Bergomi model have also been presented in El Euch et al. [2019] and Friz et al. [2022], which should lead to practically viable methods of calibration. In addition, applications to the pricing of volatility products such as VIX futures have been discussed, for example, in Jacquier et al. [2018]. Note that the rough Bergomi model is known to have the martingale property when the parameters are in certain regions (Gassiat [2019]).

Subsequently, El Euch and Rosenbaum [2019] proposed the rough Heston model, which is an extension of the Heston model (Heston [1993]), in which the stochastic integration term in the variance dynamics has been replaced with a Riemann–Liouville-type Itô integral:

$$\begin{cases} \frac{dS_t}{S_t} = \mu_t + \sqrt{V_t}\left(\rho\, dW_t^{(1)} + \sqrt{1-\rho^2}\, dW_t^{(2)}\right), \\ V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}\gamma(\theta - V_s)\, ds + \int_0^t (t-s)^{\alpha-1}\gamma\sqrt{V_s}\, dW_t^{(1)}. \end{cases}$$

This model admits a semi-analytical option pricing formula, namely the ordinary differential equation satisfied by the characteristic function of the solution is a fractional Riccati equation, which renders efficient calibration and pricing. This result parallels the fact that the corresponding ordinary differential equation in the Heston model is a Riccati equation. In addition, El Euch and Rosenbaum [2018] showed that the Markov property is recovered as in the rough Bergomi model by making the autoregressive coefficient θ time-dependent, and replication using forward variances is possible. Furthermore, Abi Jaber and El Euch [2019] showed how to approximate the rough Heston model with a family of Markovian models. This result corresponds to the fact that the rough Bergomi model is an infinite-dimensional extension of the Bergomi model. Furthermore, it is also known that the rough Heston model has a micro-foundation from a market microstructure perspective, as discussed below in section 5. The rough Heston model may also exhibit the Zumbach effect described in section

[47] Forward variances can be estimated from market prices of variance swaps of various maturities, etc.

[48] The parameters of rough Bergomi model (9) have the following correspondence with observed data: H corresponds to the curvature of the term structure of the ATM skews, $\eta\rho$ corresponds to the long-run level of ATM volatility, and ρ corresponds to the skew of volatility smiles (i.e., the cross-sections of the surface at each maturity).

3 (El Euch et al. [2020]). Finally, Fukasawa and Gatheral [2022] provided the asymptotic expansion formula of IV for the rough SABR model, which is an extension of the SABR model.

4.2.3 Joint calibration problem of SPX options' IV and VIX options' IV

Besides the representation of the negative power law, rough volatility models are expected to solve other problems that were considered difficult to solve with existing derivative pricing models. A prominent example is the joint calibration problem of SPX options' IV and VIX options' IV. Here, SPX is a U.S. equity index (the S&P 500 index) and VIX is the volatility index of SPX. VIX represents the uncertainty of the SPX index over the next 30 days, and is an example of model-free implied volatility (MFIV). MFIV is an IV that is calculated based on the fair prices of variance swaps, which can be replicated with a portfolio of vanilla options in a model-free manner (i.e., replication of variance swaps and hence MFIVs do not rely on specific pricing models such as the Black–Scholes model).[49]

The VIX is published as an index and is used for market analysis and other purposes, and futures and options written on the VIX as the underlying asset are also actively traded. Therefore, the volatility surface of the VIX is quoted in the market, and it is a challenging task to calibrate models to it and the volatility surface of the SPX simultaneously. In other words, in order to prevent arbitrage opportunities between the two classes of highly liquid instruments, an option on the SPX and options on the VIX (which can be regarded as options on options on the SPX), pricing them in a consistent and unified manner is an important issue in the U.S. equity derivatives market. Specifically, what makes this issue challenging is calibrating the parameters of the dynamics of the SPX so that the theoretical prices (IVs) of the model fit the two volatility surfaces while preserving the theoretical relationship between the SPX and the VIX. Various studies have been conducted on this problem, starting with Gatheral [2008], but most of these resorted to jump-type models, because calibration has been considered to be difficult with conventional Itô process-based models.[50] See, for example, Cont and Kokholm [2013] and Guyon [2020].

Gatheral et al. [2020] showed that this problem can be solved in a continuous stochastic process model with no jumps, which is an extension of the rough Heston model. This result is related to the phenomenon that the level of VIX is determined according to the past level of SPX, and is closely related to the Zumbach effect mentioned in section 3.

Note, however, that whether the rough volatility model is the only solution to this problem is still an open question. For example, based on about 15 years' worth of SPX and VIX IV

[49] See CBOE [2009] for detailed information on VIX.

[50] One of the reasons why this problem is so difficult is that there are numerous IVs to be fitted (i.e., the calibration problem has a huge number of constraints). Another is said to be that the short-term portion of the IV of the SPX tends to have a steeper slope in the strike direction, which requires a large value for the volatility parameter in the SPX volatility process, while the level of the IV of the VIX is lower than the level of this parameter.

data, Rømer [2022] argued that the simultaneous calibration problem can be solved with a model including a conventional Itô process that has neither roughness nor jump terms.

4.2.4 Prospects and challenges of the rough volatility models

As mentioned above, in markets where the negative power law is observed, it is at least theoretically necessary to use a rough volatility model. In addition, the rough Bergomi model and the rough Heston model can be used for hedging with forward variance as a hedge instrument, and these models have the potential to be used in practice as a powerful pricing model for derivatives. However, as discussed below, further technological development is required for practical use.

One of difficulties in the practical application of rough volatility models lies in its numerical calculation. As mentioned, in the calibration process, one needs to calculate option prices repeatedly as the expected payoff value, in order to adjust the model parameters so that the market quotes and the model prices match. In terms of the cost of calibration, the rough Heston and the SABR models are somewhat favorable because these models admit semi-analytical pricing formulae. However, in general rough volatility models, more computationally expensive calculations are required, compared with the conventional Markovian models. Horvath et al. [2021a] proposed a deep learning-based technique to solve this problem, called *deep calibration*. In addition to the calibration, to calculate prices and Greeks for (more complex) derivatives, one needs to calculate the expectations with respect to a general integrand, which is known as the *weak approximation problem*. To solve this problem, many methods and frameworks have been proposed, including a hybrid scheme (Bennedsen et al. [2017]), a tree method (Horvath et al. [2017]), and a Markovian approximation (Bayer and Breneis [2023], Cuchiero and Teichmann [2019], Hamaguchi [2023]). However, more theoretical studies as well as ones on practical implementations are required in order that rough volatility models become viable in general derivative pricing operations.[51]

The aforementioned research works suggest that rough volatility models may be applicable to the U.S. derivatives markets, where the negative power law has been observed. However, if the negative power law or some other empirical patterns indicating roughness are observed in a wider range of markets, the necessity and applicability of the rough volatility model shall increase. For example, in relation to the discussion on HV in section 3 and on MM in section 5, if assets that accommodate high-frequency trading exhibit rough volatility, it is reasonable to expect that the negative power law would be observed in the corresponding options market. In addition, since there might be some other features of the shape of volatility surfaces that suggest roughness in volatility, it is possible that roughness of volatility will be observed in various other markets. It is also noteworthy that the options markets for interest rate and foreign exchange tend to be more complex than those for equity derivatives due to

[51] Alòs et al. [2023] attempted to calculate CVA (price adjustment to reflect counterparty risk in the valuation of derivatives), which is mandatory in current derivatives practice, with a rough volatility model.

their market practices.[52] Accordingly, there are many problems specific to these markets, for example, the joint calibration of the IV of swaptions and the IV of caps.

5 Market microstructure and origins of roughness

As discussed in sections 3 and 4, an increasing number of studies have suggested that asset price volatility has roughness. Therefore, it is also important to elucidate the financial-economic mechanism that yields roughness in volatility. In this section, we briefly overview studies related to market microstructure (MM) as such an attempt. In section 5.1, we review previous studies on MM, focusing on those deriving continuous-time asset price dynamics from discrete individual market participants' trading order behavior. In section 5.2, we summarize El Euch et al. [2018], an article that models the behavior of individual market participants and the price formation mechanism, and proves under which circumstances discrete price movements converge to a continuous-time model with roughness.

5.1 Research on micro-foundations of asset price fluctuations

In theoretical analysis in the field of mathematical finance, as well as in other fields, asset prices are often modeled as continuous-time stochastic processes using Brownian motion. In actual financial markets, however, asset prices fluctuate discretely according to discrete transactions, and thus tick data (price data for each transaction) can be regarded as a discrete stochastic process. Investors also do not randomly place buy and sell orders.

MM is a research field that analyzes the impact of the institutional structure of financial markets (e.g., trading rules on exchanges) on asset price formation as well as on investors' behaviors at the micro level. One of the issues addressed in MM research is providing a *micro-foundation* to the stochastic process of asset price dynamics. As mentioned above, asset price dynamics are ultimately determined by the accumulation of tick-by-tick price movements resulting from investors' buying and selling transactions. This idea has motivated a strand of research that takes microstructural features such as investors' tick-level trading behavior (e.g., limit order placements and cancellations, market order placements) and the resulting dynamics of the *limit order book* as a starting point to derive a model of asset price dynamics over a longer time horizon. This approach has the advantage of having a micro-foundation, that is, trading behaviors of individual investors that are directly modeled. This approach is helpful to better understand the asset price formation mechanism, namely, how investors' behavioral features are related to patterns in asset price fluctuations. These points are also important from a policymaking perspective, such as the perspective taken when designing rules and regulations for exchanges.

[52]For example, in the case of foreign exchange options, the volatility is quoted with respect to *delta* and *butterfly* instead of strike. In the case of interest rate options, the volatility of swaptions (options written on a swap as the underlying asset) is quoted for the maturity of the underlying swap, maturity of the option, and strike. The volatilities of instruments such as interest rate caps and floors are also quoted.

Research on the limit order book has developed significantly since the 1990s, as the availability of individual order-level data began to improve. For example, an early study by Parlour [1998] analyzed dynamic changes in the limit order book conditions. Cont and De Larrard [2013] derived continuous-time asset price volatility models starting from the models of individual buy and sell order behaviors. They analyzed a model in which market orders, best bid limit orders, best offer limit orders, and order cancellations arrive following independent Poisson processes, and derived the probability distribution of time intervals between price changes and the probability distribution of the scaling limit of the mid-price (the average of the best bid and best ask prices). In particular, they found that the scaling limit of the mid-price converges to a Brownian motion, and its coefficient (i.e., volatility) can be expressed in terms of the arrival intensity of order placements. Lakner et al. [2016] considered a model in which the state of the limit order book is represented as a probability distribution. They assumed that this distribution fluctuates stochastically with the arrival of orders, and obtained results regarding its scaling limit.

When a market order is placed, the trade is executed by matching it to the best limit order, but if the volume of the market order exceeds that of the best limit order, then it is matched to the second best limit order and if the market order still remains, it will be matched to the third best limit order, and so on.[53] Thus, in the case of large-volume trading, execution occurs at less favorable prices than the best limit price, and also large changes in the trading price and mid-price occur. This phenomenon is called market impact. In order to mitigate such market impact, investors who wish to execute large orders may employ a trading technique called *iceberg trading*, in which orders are divided into smaller orders and executed gradually. While iceberg trading reduces market impact, it also requires a longer time to execute the entire order and thus it carries higher price risk (stemming from factors other than market impact). This trade-off between market impact and increased price risk determines the optimal execution strategy (see, for example, Almgren and Chriss [2001]). For a survey on optimal execution strategies, see Donnelly [2022], among others.[54] Iceberg trading and other investor behaviors may significantly affect the features of high-frequency price dynamics, as well as the long-run asset price dynamics (i.e., the continuous-time stochastic process obtained as a scaling limit).

5.2 Framework for analyzing roughness from MM perspective

El Euch et al. [2018] provided an example where a rough volatility model emerges as a scaling limit of one of the micro-founded models described above. In particular, El Euch et al. [2018] proved that the degree of the persistence of the impact of a price update on future price update intensity is a crucial determinant of whether the volatility of the scaling

[53]For more information on limit order book models, including those incorporating non-best limit order books, see, for example, Cont et al. [2023] and the references therein.

[54]Moreover, a vast literature studies market impact from the perspective of market liquidity. See Amihud [2002] and Goyenko et al. [2009], among others.

limit process exhibits the rough property. Specifically, they analyzed the scaling limit of a model in which tick-level upward and downward price changes are described by a point process called the *Hawkes process* and the persistence of the effect of price update on future price updates is expressed through how the *intensity* of the Hawkes process reacts to the occurrence of price update. The result of their analysis was that whether the scaling limit process poses roughness depends on the parameters related to the intensity of the Hawkes process. Subsequent studies have discussed generalizations of the El Euch et al. [2018] model, as well as the applications of the model to the analysis of the rough Heston model.

5.2.1 Hawkes process

To formally define the Hawkes process, we first describe the Poisson distribution and Poisson process. The Poisson distribution is a discrete probability distribution that expresses the number of occurrences of certain contingent events. It has a parameter μ that controls the likelihood of the occurrence of the events (i.e., the average number of occurrences over a unit time interval). Formally, a discrete random variable X follows a Poisson distribution with parameter μ if it satisfies the following equation:

$$P(X = k) = \frac{\mu^k}{k!} e^{-\mu}.$$

Note that both the expected value and variance of the Poisson distribution equal μ (i.e., $E[X] = Var[X] = \mu$).

The Poisson process L_t is the most basic point process that counts the occurrence of certain contingent events whose occurrence follows a Poisson distribution. Specifically, the number of occurrences of events between time t and $t + s$, $L_{t+s} - L_t$, follows a Poisson distribution with parameter $\mu = \lambda s$. Namely, the following holds:

$$P(L_{t+s} - L_t = k) = \frac{(\lambda s)^k}{k!} e^{-\lambda s},$$

where λ is a parameter called the intensity of the Poisson process. As can be seen from the fact that the average number of events over time duration s is λs, the intensity λ represents the instantaneous likelihood of the occurrence of an event, that is, $P(L_{t+dt} - L_t = 1) = \lambda dt$ holds.[55]

As such, the intensity of the (stationary) Poisson process is constant. However, there exist a variety of real-life contingent events, where the occurrence of an event stimulates following events, such as aftershocks following a main earthquake, chain reaction bankruptcies after the bankruptcy of a large company, or sell-off after a sharp market decline. In other words, the occurrence of an event increases the intensity of following events in these cases, and this property is called the self-exciting property. The Hawkes process is an extension of the Poisson process that has the self-exciting property. The intensity λ_t of the Hawkes process

[55] Mathematically, a Poisson process is defined to be an independent increments process that satisfies $P(L_{t+dt} - L_t = 1) = \lambda dt$ and $P(L_{t+dt} - L_t \geq 2) = 0$.

N_t at a certain point in time depends on the number of the events that have occurred up to that point (that is, N_t). Specifically, the intensity of the Hawkes process is given as the following equation:

$$\lambda_t = \int_0^t h(t-s)\, dN_s,$$

where the function h is called a kernel function of the Hawkes process. For example, if the kernel function is positive (i.e., $h(\cdot) > 0$), then the intensity will increase with the occurrence of events and the process exhibits the self-exciting property. If the kernel function h is a positive-valued decreasing function, the effect of the occurrence of past events on the intensity will decay over time.

5.2.2 Framework of El Euch et al. [2018]

El Euch et al. [2018] modeled tick-by-tick price movements using a Hawkes process and proved that the convergence property of the intensity determines whether roughness arises in the volatility of the price process obtained as a scaling limit. Below, we give a brief explanation of this article.

Let N_t^+ and N_t^- be the numbers of price increases and decreases, respectively, and (N_t^+, N_t^-) be a two-dimensional Hawkes process. Namely, let λ^+ and λ_t^- be the intensities of N_t^+ and N_t^-, respectively, given by the following equation:

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \begin{pmatrix} \nu^+ \\ \nu^- \end{pmatrix} + \int_0^t \begin{pmatrix} \phi_1(t-s) & \phi_2(t-s) \\ \phi_3(t-s) & \phi_4(t-s) \end{pmatrix} \begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix}, \tag{10}$$

where the kernel functions ϕ_1, ϕ_2, ϕ_3, ϕ_4 represent how past price increases and decreases affect the intensity of price increases and decreases. For example, the first row of (10) is the intensity of the price increase and is written as

$$\lambda_t^+ = \nu^+ + \int_0^t \phi_1(t-s)\, dN_s^+ + \int_0^t \phi_2(t-s)\, dN_s^-.$$

The second term on the right-hand side represents the impact of past price increases on the intensity of current price increases, and the third term on the right-hand side represents the impact of past price decreases on the intensity of current price increases. Similarly, the intensity of price decrease is affected by past price increases and decreases. Thus, the model represents the so-called mutually exciting property, with which price increases and price decreases interact with each other to affect the intensity of future price changes.

Under this framework, the article puts additional conditions on the self-exciting parameters and discusses whether roughness arises in the stochastic process followed by the scaling limit P_t formed under high-frequency trading.[56] As a result, if the effect of price changes on

[56]The scaling limit is the price $P_t = \lim_{T\to\infty} \frac{N_{tT}^+ - N_{tT}^-}{T}$ formed under a hypothetical trade in infinitesimal time.

the intensity lingers over a long period of time, that is, in the case that the kernel function,

$$\Phi(u) = \begin{pmatrix} \phi_1(u) & \phi_2(u) \\ \phi_3(u) & \phi_4(u) \end{pmatrix},$$

decays slowly as $u \to \infty$, roughness arises.[57]

El Euch et al. [2018] interpreted this condition for roughness as the market trading behavior of iceberg trading described above, where a large order is split into consecutive smaller trade orders over a longer period of time so that an order is followed by many other orders. However, this is one of the interpretations of the convergence condition on the intensity of the Hawkes process described above; it has not been excluded that other patterns of investors' behaviors may result in the same mathematical condition. Therefore, the result by El Euch et al. [2018] does not immediately lead to the conclusion that iceberg trading is the source of roughness in high-frequency trading.

5.2.3 Prospects for rough volatility research from MM perspective

Subsequent studies have extended the framework of El Euch et al. [2018]. For example, Jusselin and Rosenbaum [2020] clarified the correspondence between the degree of impact of a transaction on market prices and roughness (i.e., Hurst parameter H) in a more general setting of the market. In addition, Rosenbaum and Tomas [2021] extended the discussion of two-dimensional Hawkes processes to multidimensional Hawkes processes and presented a framework for analyzing micro-structures in markets where multiple assets are traded. Further research from the MM perspective, including more realistic modeling of investors' behaviors and empirical analysis using real data, would be important to elucidate the source of the roughness of volatility.

6 Concluding remarks

In this paper, we have surveyed empirical and theoretical studies on rough volatility and discussed its potential applications from various perspectives. In particular, as evidence for the roughness of volatility, we have introduced statistical observations on time-series data for realized volatility (section 3) and the shape of IV observed in derivatives markets (section 4).

[57]For example, if there exists a constant $1/2 < \alpha < 1$ such that $\alpha x^{\alpha} \int_x^{\infty} \phi_1(s)\, ds \to C$ $(x \to \infty)$, then the scaling limit converges to the rough Heston model. Here, α appearing in the convergence condition corresponds to the Hurst index of the rough Heston model obtained as the scale limit through $H = \alpha - 1/2$. The technical gist of this result is as follows. Under high-frequency trading, the kernel function satisfies certain integrability conditions (such as $\int_0^{\infty} \phi_1(s)\, ds$ converging in an appropriate sense, a condition called being a *nearly unstable* Hawkes process). Under this condition, they discussed the convergence of the eigenvalue of $\Phi(u) = \begin{pmatrix} \phi_1(u) & \phi_2(u) \\ \phi_3(u) & \phi_4(u) \end{pmatrix}$. The degree of decay of these eigenvalues determines whether the convergent scale limit process has roughness.

Both of these observations can be expressed consistently using fractional Brownian motion, a model that has a rough feature.

Furthermore, in section 4, we have explained an important theoretical result that indicates the necessity of rough volatility models. Namely, for derivatives markets where roughness is observed, a pricing model with appropriate roughness is necessary because otherwise arbitrage opportunities arise. Section 5 introduced a study from the MM research strand that attempted to shed light on the mechanism by which roughness arises. In addition, as applications, roughness could improve volatility forecasting accuracy (section 3) and derivative pricing models (section 4). As such, the necessity and usefulness of roughness have been discussed extensively in various fields, including financial econometrics and mathematical finance, which originally developed around different issues. Rough volatility has become a new trend in volatility research.

It should be noted, however, that although a growing number of studies have suggested the existence of roughness in volatility, there is an ongoing discussion as to whether rough volatility modeling is really necessary. Most of the studies to date have suggested that the assumption of rough volatility is just a sufficient condition to explain empirical observations consistently. In addition, roughness has only been empirically confirmed in a few markets, such as the U.S. equity market. Therefore, it is important to deepen our understanding of whether the roughness of volatility is a universal property that appears regardless of asset class, as well as other issues. It is also important to uncover financial economic mechanisms behind the stylized features revealed by the existing empirical studies, especially what kind of trading behaviors of market participants give rise to the roughness. Therefore, further research from the perspective of MM is desirable, as described in section 5. Even if the roughness of volatility is confirmed, there are various challenges for application. For example, it is debatable whether the models based on fractional Brownian motion and Riemann–Liouville-type Itô integral formulations presented in this paper are desirable in terms of representability and implementability. Various other issues remain to be addressed, including parameter estimation for time-series models, stabilization of numerical methods for derivative pricing models, and analysis of hedging strategies.

Rough volatility research is expected to develop further, expanding its related fields in both practical and academic directions. For example, application to high-frequency (algorithmic) trading (Guasoni et al. [2021]), integration with machine learning and financial time-series data generation (Rosenbaum and Zhang [2022], Horvath et al. [2021b]), and application to broad risk management have already begun. It is expected that research will continue to progress, influenced by changes in the issues of interest in financial practice, while also intersecting with the fields of artificial intelligence, economics, information science, and mathematical science. Therefore, it is important to keep abreast of the leading-edge discussions.

References

Abi Jaber, E. and El Euch, O., "Multifactor Approximation of Rough Volatility Models", *SIAM Journal on Financial Mathematics*, 10(2), pp.309–349, 2019.

Aït-Sahalia, Y. and Mykland, P. A., "Estimating Volatility in the Presence of Market Microstructure Noise: A Review of the Theory and Practical Considerations", In T. Mikosch, J.-P. Kreiß, R. A. Davisand , and T. G. Andersen (Eds.), *Handbook of Financial Time Series* pp. 577–598, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

Almgren, R. and Chriss, N., "Optimal Execution of Portfolio Transactions", *The Journal of Risk*, 3(2), pp.5–39, 2001.

Alòs, E., Antonelli, F., Ramponi, A., and Scarlatti, S., "CVA in fractional and rough volatility models", *Applied Mathematics and Computation*, 442, pp.127715, 2023.

Alòs, E. and León, J. A., "An Intuitive Introduction to Fractional and Rough Volatilities", *Mathematics*, 9(9), 2021.

Alòs, E., León, J. A., and Vives, J., "On the Short-Time Behavior of the Implied Volatility for Jump-Diffusion Models with Stochastic Volatility", *Finance and Stochastics*, 11(4), pp.571–589, 2007.

Amihud, Y., "Illiquidity and Stock Returns: Cross-section and Time-Series Effects", *Journal of Financial Markets*, 5, pp.31–56, 2002.

Andersen, T. G., Bollerslev, T., and Diebold, F. X., "Roughing It up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility", *The Review of Economics and Statistics*, 89(4), pp.701–720, 2007.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P., "The Distribution of Realized Exchange Rate Volatility", *Journal of the American Statistical Association*, 96(453), pp.42–55, 2001.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P., "Modeling and Forecasting Realized Volatility", *Econometrica*, 71(2), pp.579–625, 2003.

Andersen, T. G., Dobrev, D., and Schaumburg, E., "Jump-Robust Volatility Estimation Using Nearest Neighbor Truncation", *Journal of Econometrics*, 169(1), pp.75–93, 2012.

Avramov, D., Chordia, T., and Goyal, A., "The Impact of Trades on Daily Volatility", *Review of Financial Studies*, 19(4), pp.1241–1277, 2006.

Bachelier, L., "Théorie de la spéculation", *Annales scientifiques de l'École Normale Supérieure*, 17, pp.21–86, 1900.

Baillie, R. T., "Long Memory Processes and Fractional Integration in Econometrics", *Journal of Econometrics*, 73(1), pp.5–59, 1996.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O., "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 74(1), pp.3–30, 1996.

Baker, S. R., Bloom, N., and Davis, S. J., "Measuring Economic Policy Uncertainty", *Quarterly Journal of Economics*, 131(4), pp.1593–1636, 2016.

Bandi, F. M. and Russell, J. R., "Microstructure Noise, Realized Variance, and Optimal Sampling", *Review of Economic Studies*, 75(2), pp.339–369, 2008.

Barndorff-Nielsen, O. E., "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation", *Journal of Financial Econometrics*, 4(1), pp.1–30, 2005.

Barndorff-Nielsen, O. E. and Schmiegel, J., "Ambit Processes; with Applications to Turbulence and Tumour Growth", In F. E. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendaland , and T. Zhang (Eds.), *Stochastic Analysis and Applications* pp. 93–124, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

Barndorff-Nielsen, O. E. and Schmiegel, J., "Brownian Semistationary Processes and Volatility/Intermittency", In H. Albrecher, W. J. Runggaldierand , and W. Schachermayer (Eds.), *Advanced Financial Modelling* pp. 1–26, Walter de Gruyter, 2009.

Barndorff-Nielsen, O. E. and Shephard, N., "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models", *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(2), pp.253–280, 2002.

Barndorff-Nielsen, O. E. and Shephard, N., "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics", *Econometrica*, 72(3), pp.885–925, 2004.

Bates, D. S., "Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options", *The Review of Financial Studies*, 9(1), pp.69–107, 1996.

Bayer, C. and Breneis, S., "Markovian approximations of stochastic Volterra equations with the fractional kernel", *Quantitative Finance*, 23(1), pp.53–70, 2023.

Bayer, C., Friz, P., and Gatheral, J., "Pricing under Rough Volatility", *Quantitative Finance*, 16(6), pp.887–904, 2016.

Bayer, C., Friz, P. K., Gassiat, P., Martin, J., and Stemper, B., "A regularity structure for rough volatility", *Mathematical Finance*, 30(3), pp.782–832, 2020.

Bekaert, G. and Hoerova, M., "The VIX, the Variance Premium and Stock Market Volatility", *Journal of Econometrics*, 183(2), pp.181–192, 2014.

Bennedsen, M., Lunde, A., and Pakkanen, M. S., "Hybrid scheme for Brownian semistationary processes", *Finance and Stochastics*, 21, pp.931–965, 2017.

Bennedsen, M., Lunde, A., and Pakkanen, M. S., "Decoupling the Short- and Long-Term Behavior of Stochastic Volatility", *Journal of Financial Econometrics*, 20(5), pp.961–1006, 2021.

Bergomi, L., "Smile dynamics I", *Available at SSRN 1493294*, 2004.

Best, M. J. and Grauer, R. R., "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results", *Review of Financial Studies*, 4(2), pp.315–342, 1991.

Björk, T., *Arbitrage theory in continuous time*, Oxford university press, 2009.

Black, F., "Studies of Stock Price Volatility Changes", In *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section* pp. 177–181, American Statistical Association, 1976.

Black, F. and Scholes, M., "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81(3), pp.637–654, 1973.

Blanc, P., Donier, J., and Bouchaud, J.-P., "Quadratic Hawkes Processes for Financial Prices", *Quantitative Finance*, 17(2), pp.171–188, 2017.

Bloom, N., "The Impact of Uncertainty Shocks", *Econometrica*, 77(3), pp.623–685, 2009.

Bloom, N., Bond, S., and van Reenen, J., "Uncertainty and Investment Dynamics", *The Review of Economic Studie*, 74(2), pp.391–415, 2007.

Bolko, A. E., Christensen, K., Pakkanen, M. S., and Veliyev, B., "A GMM Approach to Estimate the Roughness of Stochastic Volatility", *Journal of Econometrics*, 235(2), pp.745–778, 2023.

Bollerslev, T., "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31(3), pp.307–327, 1986.

Bollerslev, T. and Mikkelsen, H. O., "Modeling and Pricing Long Memory in Stock Market Volatility", *Journal of Econometrics*, 73(1), pp.151–184, 1996.

Breidt, F., Crato, N., and De Lima, P., "The Detection and Estimation of Long Memory in Stochastic Volatility", *Journal of Econometrics*, 83(1-2), pp.325–348, 1998.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C., *The Econometrics of Financial Markets*, Princeton, N.J: Princeton University Press, 1997.

Carr, P. and Wu, L., "The finite moment log stable process and option pricing", *The Journal of Finance*, 58(2), pp.753–777, 2003.

CBOE, "The CBOE Volatility Index—VIX", *Chicago Board Options Exchange, White paper*, 2009.

Cheridito, P., Kawaguchi, H., and Maejima, M., "Fractional Ornstein-Uhlenbeck Processes", *Electronic Journal of Probability*, 8, 2003.

Christie, A., "The Stochastic Behavior of Common Stock Variances Value, Leverage and Interest Rate Effects", *Journal of Financial Economics*, 10(4), pp.407–432, 1982.

Comte, F., Coutin, L., and Renault, E., "Affine Fractional Stochastic Volatility Models", *Annals of Finance*, 8(2-3), pp.337–378, 2012.

Comte, F. and Renault, E., "Long Memory Continuous Time Models", *Journal of Econometrics*, 73(1), pp.101–149, 1996.

Comte, F. and Renault, E., "Long Memory in Continuous-Time Stochastic Volatility Models", *Mathematical Finance*, 8(4), pp.291–323, 1998.

Cont, R. and Das, P., "Rough Volatility: Fact or Artefact?", 2023.

Cont, R. and De Larrard, A., "Price Dynamics in a Markovian Limit Order Market", *SIAM Journal on Financial Mathematics*, 4(1), pp.1–25, 2013.

Cont, R., Degond, P., and Xuan, L., "A Mathematical Framework for Modelling Order Book Dynamics", 2023.

Cont, R. and Kokholm, T., "A consistent pricing model for index options and volatility derivatives", *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(2), pp.248–274, 2013.

Corsi, F., "A Simple Approximate Long-Memory Model of Realized Volatility", *Journal of Financial Econometrics*, 7(2), pp.174–196, 2008.

Cuchiero, C. and Teichmann, J., "Markovian lifts of positive semidefinite affine Volterra-type processes", *Decisions in Economics and Finance*, 42, pp.407–448, 2019.

Damian, C. and Frey, R., "Detecting Rough Volatility: A Filtering Approach", 2023.

Delemotte, J., Marco, S. D., and Segonne, F., "Yet Another Analysis of the SP500 At-The-Money Skew: Crossover of Different Power-Law Behaviours", *Available at SSRN 4428407*, 2023.

Dew-Becker, I., Giglio, S., and Kelly, B., "Hedging Macroeconomic and Financial Uncertainty and Volatility", *Journal of Financial Economics*, 142(1), pp.23–45, 2021.

Di Nunno, G., Kubilius, K., Mishura, Y., and Yurchenko-Tytarenko, A., "From Constant to Rough: A Survey of Continuous Volatility Modeling", *Mathematics*, 11(19), 2023.

Diebold, F. X. and Inoue, A., "Long Memory and Regime Switching", *Journal of Econometrics*, 105, pp.131–159, 2001.

Diebold, F. X. and Mariano, R. S., "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, 20(1), pp.134–144, 2002.

Ding, Z., Granger, C. W. J., and Engle, F., "A Long Memory Property of Stock Market Returns and a New Model", *Journal of Empirical Finance*, 1, pp.83–106, 1993.

Donnelly, R., "Optimal Execution: A Review", *Applied Mathematical Finance*, 29(3), pp.181–212, 2022.

Dupire, B., "Pricing with a smile", *Risk*, 7(1), pp.18–20, 1994.

El Euch, O., Fukasawa, M., Gatheral, J., and Rosenbaum, M., "Short-term at-the-money asymptotics under stochastic volatility models", *SIAM Journal on Financial Mathematics*, 10(2), pp.491–511, 2019.

El Euch, O., Fukasawa, M., and Rosenbaum, M., "The Microstructural Foundations of Leverage Effect and Rough Volatility", *Finance and Stochastics*, 22(2), pp.241–280, 2018.

El Euch, O., Gatheral, J., Radoičić, R., and Rosenbaum, M., "The Zumbach Effect under Rough Heston", *Quantitative Finance*, 20(2), pp.235–241, 2020.

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

El Euch, O. and Rosenbaum, M., "The Characteristic Function of Rough Heston Models", *Mathematical Finance*, 29(1), pp.3–38, 2019.

Engle, R. F., "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50(4), pp.987–1007, 1982.

Fama, E. F., "The Behavior of Stock-Market Prices", *The Journal of Business*, 38(1), pp.34–105, 1965.

Fouque, J.-P., Papanicolaou, G., Sircar, R., and Solna, K., "Multiscale stochastic volatility asymptotics", *Multiscale Modeling & Simulation*, 2(1), pp.22–42, 2003.

Friz, P. K., Gassiat, P., and Pigato, P., "Short-dated smile under rough volatility: asymptotics and numerics", *Quantitative Finance*, 22(3), pp.463–480, 2022.

Fukasawa, M., "Asymptotic Analysis for Stochastic Volatility: Martingale Expansion", *Finance and Stochastics*, 15(4), pp.635–654, 2011.

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

Fukasawa, M., “Volatility Has to Be Rough”, *Quantitative Finance*, 21(1), pp.1–8, 2021.

Fukasawa, M. and Gatheral, J., “A Rough SABR Formula”, *Frontiers of Mathematical Finance*, 1(1), pp.81–97, 2022.

Fukasawa, M., Horvath, B., and Tankov, P., “Hedging under rough volatility”, *arXiv preprint arXiv:2105.04073*, 2021.

Fukasawa, M., Takabatake, T., and Westphal, R., “Consistent Estimation for Fractional Stochastic Volatility Model under High-frequency Asymptotics”, *Mathematical Finance*, 32(4), pp.1086–1132, 2022.

Funahashi, H. and Kijima, M., “Does the Hurst index matter for option prices under fractional volatility?”, *Annals of Finance*, 13, pp.55–74, 2017a.

Funahashi, H. and Kijima, M., “A solution to the time-scale fractional puzzle in the implied volatility”, *Fractal and fractional*, 1(1), pp.14, 2017b.

Gassiat, P., “On the martingale property in the rough Bergomi model”, 2019.

Gatheral, J., “Consistent modeling of SPX and VIX options”, In *Bachelier congress*, volume 37 pp. 39–51, 2008.

Gatheral, J., Jaisson, T., and Rosenbaum, M., “Volatility Is Rough”, *Quantitative Finance*, 18(6), pp.933–949, 2018.

Gatheral, J., Jusselin, P., and Rosenbaum, M., “The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem”, *arXiv preprint arXiv:2001.01789*, 2020.

Gençay, R., Ballocchi, G., Dacorog, M., Olsen, R., and Pictet, O., “Real-Time Trading Models and the Statistical Properties of Foreign Exchange Rates”, *International Economic Review*, 43(2), pp.463–491, 2002.

Glosten, L. R., Jagannathan, R., and Runkle, D. E., “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”, *The Journal of Finance*, 48(5), pp.1779–1801, 1993.

Goyenko, R. Y., Holden, C. W., and Trzcinka, C. A., “Do Liquidity Measures Measure Liquidity?”, *Journal of Financial Economics*, 92(2), pp.153–181, 2009.

Granger, C. W., “The Typical Spectral Shape of an Economic Variable”, *Econometrica*, 34(1), pp.150–161, 1966.

Granger, C. W., “Long Memory Relationships and the Aggregation of Dynamic Models”, *Journal of Econometrics*, 14, pp.227–238, 1980.

Granger, C. W. and Hyung, N., "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns", *Journal of Empirical Finance*, 11(3), pp.399–421, 2004.

Granger, C. W. J. and Joyeux, R., "An Introduction to Long-Memory Time Series Models and Fractional Differencing", *Journal of Time Series Analysis*, 1(1), pp.15–29, 1980.

Guasoni, P., Mishura, Y., and Rásonyi, M., "High-Frequency Trading with Fractional Brownian Motion", *Finance and Stochastics*, 25(2), pp.277–310, 2021.

Guyon, J., "The joint S&P 500/VIX smile calibration puzzle solved", *Risk, April*, 2020.

Guyon, J. and El Amrani, M., "Does the Term-Structure of Equity At-the-Money Skew Really Follow a Power Law?", *Available at SSRN 4174538*, 2022.

Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E., "Managing smile risk", *The Best of Wilmott*, 1, pp.249–296, 2002.

Hamaguchi, Y., "Markovian lifting and asymptotic log-Harnack inequality for stochastic Volterra integral equations", 2023.

Hamilton, J. D., *Time Series Analysis*, Princeton, NJ: Princeton Univ. Press, 1994.

Hansen, P. R. and Lunde, A., "Realized Variance and Market Microstructure Noise", *Journal of Business & Economic Statistics*, 24(2), pp.127–161, 2006.

Hansen, P. R., Lunde, A., and Nason, J. M., "The Model Confidence Set", *Econometrica*, 79(2), pp.453–497, 2011.

Haubrich, J. G. and Lo, A. W., "The Sources and Nature of Long-Term Memory in the Business Cycle", 1991.

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

Horvath, B., Jacquier, A., and Muguruza, A., "Functional central limit theorems for rough volatility", *arXiv preprint arXiv:1711.03078*, 2017.

Horvath, B., Muguruza, A., and Tomas, M., "Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models", *Quantitative Finance*, 21(1), pp.11–27, 2021a.

Horvath, B., Teichmann, J., and Zuric, Z., 2021b, "Deep Hedging under Rough Volatility".

Hosking, J. R. M., "Fractional Differencing", *Biometrika*, 68(1), pp.165–176, 1981.

Hurst, H. E., "Long-Term Storage Capacity of Reservoirs", *Transactions of the American Society of Civil Engineers*, 116(1), pp.770–799, 1951.

Jacquier, A., Martini, C., and Muguruza, A., "On VIX futures in the rough Bergomi model", *Quantitative Finance*, 18(1), pp.45–61, 2018.

Jagannathan, R. and Ma, T., "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps", *Journal of Finance*, 58(4), pp.1651–1683, 2003.

Jusselin, P. and Rosenbaum, M., "No-arbitrage implies power-law market impact and rough volatility", *Mathematical Finance*, 30(4), pp.1309–1336, 2020.

Karatzas, I. and Shreve, S. E., *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*, New York, NY: Springer New York, 1998.

Lakner, P., Reed, J., and Stoikov, S., "High Frequency Asymptotics for the Limit Order Book", *Market Microstructure and Liquidity*, 02(01), pp.1650004, 2016.

Lee, R. W., "Implied volatility: Statics, dynamics, and probabilistic interpretation", *Recent advances in applied probability*, pp. 241–268, 2005.

Levy, P. B., "Random functions:general theory with special reference to Laplacian random functions", In *University of California Publications in Statistics* pp. 331–390, 1953.

Lintner, J., "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", *The Review of Economics and Statistics*, 47(1), pp.13–37, 1965.

Liu, L. Y., Patton, A. J., and Sheppard, K., "Does Anything Beat 5-Minute RV? A Comparison of Realized Measures across Multiple Asset Classes", *Journal of Econometrics*, 187(1), pp.293–311, 2015.

Lynch, P. E. and Zumbach, G. O., "Market Heterogeneities and the Causal Structure of Volatility", *Quantitative Finance*, 3(4), pp.320–331, 2003.

Lyons, T. J., "Differential equations driven by rough signals", *Revista Matemática Iberoamericana*, 14(2), pp.215–310, 1998.

Mandelbrot, B., "The Variation of Certain Speculative Prices", *The Journal of Business*, 36(4), pp.394–419, 1963.

Mandelbrot, B. B. and Van Ness, J. W., "Fractional Brownian motions, fractional noises and applications", *SIAM review*, 10(4), pp.422–437, 1968.

Markowitz, H., "Portfolio Selection", *Journal of Finance*, 7(1), pp.77–91, 1952.

Martin, I., "What Is the Expected Return on the Market?", *Quarterly Journal of Economics*, 132(1), pp.367–433, 2017.

Merton, R. C., "Theory of Rational Option Pricing", *The Bell Journal of Economics and Management Science*, 4(1), pp.141–183, 1973.

Merton, R. C., “On Estimating the Expected Return on the Market: An Exploratory Investigation”, *Journal of Financial Economics*, 8(4), pp.323–361, 1980.

Michaud, R. O., “The Markowitz Optimization Enigma: Is ’Optimized’ Optimal?”, *Financial Analysts Journal*, 45(1), pp.31–42, 1989.

Mossin, J., “Equilibrium in a Capital Asset Market”, *Econometrica*, 34(4), pp.768, 1966.

Nagakura, D. and Watanabe, T., “A State Space Approach to Estimating the Integrated Variance under the Existence of Market Microstructure Noise”, *Journal of Financial Econometrics*, 13(1), pp.45–82, 2015.

Nelson, D. B., “Conditional Heteroskedasticity in Asset Returns: A New Approach”, *Econometrica*, 59(2), pp.347, 1991.

Nerlove, M., “Spectral Analysis of Seasonal Adjustment Procedures”, *Econometrica*, 32(3), pp.241, 1964.

Nualart, D., *The Malliavin Calculus and Related Topics*, Probability and Its Applications, Berlin ; New York: Springer, 2nd ed edition, 2006.

Osajima, Y., “The asymptotic expansion formula of implied volatility for dynamic SABR model and FX hybrid model”, *Available at SSRN 965265*, 2007.

Parlour, C. A., “Price Dynamics in Limit Order Markets”, *Review of Financial Studies*, 11(4), pp.789–816, 1998.

Poon, S.-H. and Granger, C. W. J., “Forecasting Volatility in Financial Markets: A Review”, *Journal of Economic Literature*, 61, pp.478–539, 2003.

Rogers, L.C.G., “Things We Think We Know”, In D. Gershon, A. Lipton, M. Rosenbaum, and Z. Wiener (Eds.), Options—45 Years since the Publication of the Black–Scholes–Merton Model pp. 173–184, World Scientific, 2023.

Rømer, S. E., “Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets”, *Quantitative Finance*, 22(10), pp.1805–1838, 2022.

Rosenbaum, M. and Tomas, M., “From microscopic price dynamics to multidimensional rough volatility models”, *Advances in Applied Probability*, 53(2), pp.425–462, 2021.

Rosenbaum, M. and Zhang, J., “On the universality of the volatility formation process: when machine learning and rough volatility agree”, *arXiv preprint arXiv:2206.14114*, 2022.

Sharpe, W. F., “Mutual Fund Performance”, *The Journal of Business*, 39(1, Part 2: Supplement on Security Prices), pp.119–138, 1966.

Sowell, F., “Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models”, *Journal of Econometrics*, 53, pp.165–188, 1992.

Takahashi, M., Omori, Y., and Watanabe, T., *Stochastic Volatility and Realized Stochastic Volatility Models*, SpringerBriefs in Statistics, Singapore: Springer Nature Singapore, 2023.

Ubukata, M. and Watanabe, T., "Pricing Nikkei 225 Options Using Realized Volatility: Pricing Nikkei 225 Options Using Realized Volatility", *Japanese Economic Review*, 65(4), pp.431–467, 2014.

Wang, X., Xiao, W., and Yu, J., "Modeling and Forecasting Realized Volatility with the Fractional Ornstein–Uhlenbeck Process", *Journal of Econometrics*, 232(2), pp.389–415, 2023a.

Wang, X., Yu, J., and Zhang, C., 2023b, "On the Optimal Forecast with the Fractional Brownian Motion".

Wu, G., "The Determinants of Asymmetric Volatility", *Review of Financial Studies*, 14(3), pp.837–859, 2001.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y., "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data", *Journal of the American Statistical Association*, 100(472), pp.1394–1411, 2005.

Zhou, B., "High-Frequency Data and Volatility in Foreign-Exchange Rates", *Journal of Business & Economic Statistics*, 14(1), pp.45, 1996.

Zumbach, G., "Volatility Processes and Volatility Forecast with Long Memory", *Quantitative Finance*, 4(1), pp.70–86, 2004.

Zumbach, G., "Time Reversal Invariance in Finance", *Quantitative Finance*, 9(5), pp.505–515, 2009.

Zumbach, G. and Lynch, P., "Heterogeneous Volatility Cascade in Financial Markets", *Physica A*, 298, pp.521–529, 2001.